THE INFLUENCE OF THE LEVELS OF FIDELITY OF IMPLEMENTATION OF THE REAPS MODEL ON STUDENTS' CREATIVITY IN SCIENCE

by

Fahad Suliman Alfaiz

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF DISABILITY AND PSYCHOEDUCATIONAL STUDIES

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

WITH A MAJOR IN SPECIAL EDUCATION

In the Graduate College

THE UNIVERSITY OF ARIZONA

2019

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by ***Fahad Suliman Alfaiz***, titled ***The Influence of the Levels of Fidelity of Implementation of the REAPS Model on Students' Creativity in Science*** and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

C. June Maker
C. June Maker Date: ***3/18/2019***

Carl J. Liaupsin Date: ***3/18/2019***

Kathi Carter
Katherine J. Carter Date: ***3/18/2019***

L. Penny Rosenblum Date: ***3/18/2019***

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

We hereby certify that we have read this dissertation prepared under our direction and recommend that it be accepted as fulfilling the dissertation requirement.

C. June Maker
C. June Maker Date: 4/24/19
Professor Emerita
Disability & Psychoeducational Studies

Carl J. Liaupsin Date: 4/25/19
Professor
Department Head of Disabilities and Psychoeducational Studies

STATEMENT BY AUTHOR

SIGNED: Fahad Suliman Alfaiz

ACKNOWLEDGEMENTS

I would like to thank my academic advisor, Dr. C. June Maker, for her guidance, support, and encouragement during my PhD journey. I was fortunate to be one of her final graduate students. Under her supervision, I have had many opportunities to participate in her DISCOVER projects and other research and have learned many skills that will help me in my future career as a researcher.

A special thanks goes to Dr. Katherine J. Carter, my minor advisor, for her commitment to providing insightful feedback during my coursework and dissertation process. I am also grateful to my committee members, Dr. Carl Liaupsin, a co-chairperson of my committee and the Department Head of Disability & Psychoeducational Studies, and Dr. L. Penny Rosenblum, for their continual flexibility, advice, and support. In addition, I would like to thank all my professors for the knowledge that they shared and the skills that they helped me acquire.

I would like to thank Dr. Heidi Legg Burross, faculty member in the Educational Psychology department, as well as Dr. Mohammad Torabi and Dr. Mark Borgstrom, Research Computing Specialists and Statistical Consultants, for their advice during the statistical analysis stage of my project.

Finally, I would like to thank all of the DISCOVER research team, Dr. Robert Zimmerman, Randy Pease, Dr. Sonmi Jo, Dr. Abdulnasser Alhusaini, and I-Chen Wu, for their collaboration and advice.

DEDICATION

This dissertation would not have been accomplished without the great support I received from my family during my doctoral program. The prayers and wishes that I heard from my mother, Aljohara, and my father, Suliman, every time I talked to them had a positive impact on my motivation to complete this work. A special thank you goes to my wife, Rasha, for her undying support, understating, and patience during the years that I spent on my degree. I owe her the time that she dedicated to taking care of our beautiful daughter and son, Haya and Mohammed, while I was busy writing my dissertation. Finally, I would like to thank my brothers and sisters for their prayers.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Developing students' creative problem solving skills should be one of the top goals in general education. One teaching model that was developed to increase students' creative problem solving abilities is the Real Engagement in Active Problem Solving (REAPS) model. This model was used as an intervention in classrooms in many studies. However, the researchers of these studies did not examine the relationship between the levels of implementation of this model and its effect on students' creativity in science. In the current study, the author aimed to determine the influence of the level of fidelity of implementation (FOI) of the REAPS model on creative problem solving in science. Specifically, the purpose was to examine the relationships between changes in students' growth in creative problem solving in science and the level of FOI of the REAPS model.

A quasi-experiment design was used, and the study took place in one elementary public school in Australia for two years. The total number of students who participated was 317. In Year 1, the number of classrooms was 17, and students were in grade 2 to 5. In Year 2, the number of classes was 16, and students were in grades 3 through 6. The instrument used to collect data was the Test of Creative Problem Solving in Science (TCPS-S). Students took a pretest prior to exposure to REAPS in Year 1, and they took a posttest at the end of Year 2.

One-Way Repeated Measures Multivariate Analysis of Variance (MANOVA) was used to answer the three research questions. Five dependent variables were used in Questions 1 and 2: the total scores on TCPS-S, fluency, flexibility, elaboration, and originality. Grade levels were used as an independent variable for Question 1, and levels of FOI were used as an independent variable for Question 2. For Question 3, three dependent variables were used: Finding Problems,

Generating Detailed Solutions, and Classifying Elements with one independent variable: levels of FOI.

In general, students showed growth in their science creativity scores from pretest to posttest. Lower grade students showed growth on originality, which was statistically significant, compared with students in upper grades (i.e., grades 4 to 6), while the upper grade students showed growth on the total scores on TCPS-S, fluency, flexibility, and elaboration, which was statistically significant. Students who were with high implementers for two consecutive years or with high implementers in the second year showed growth on flexibility and elaboration, which was statistically significant, compared with students who were with low implementers for two years or with low implementers in the second year. No differences were found among the levels of FOI on the total scores on TCPS-S, fluency, and originality. Students who were with high implementers for two consecutive years or with high implementers in the second year showed growth in their scores in two factors of TCPS-S: Generating Detailed Solutions and Classifying Elements. However, no differences were found in Finding Problems factor among the levels of FOI. Future research is recommended to examine the effects of the use of REAPS on students' creativity in different disciplines.

Keywords: creativity in science, fidelity of implementation, problem solving, REAPS, TCPS-S, fluency, flexibility, elaboration, originality

CHAPTER I: INTRODUCTION

Developing students' creative problem solving abilities must be a high priority within the education system of any country that seeks to maintain a healthy environment for its people. The world is becoming increasingly more complex, and the challenges which governments and people face have increased dramatically due to various reasons, such as the growth of global population, food shortages and water pollution, especially in developing countries, and the negative impacts of rapid industrial development.

In schools, the teaching approaches that students receive in classrooms are varied. Some of these approaches are focused on memory, with students recalling information without connecting it to real-world problems, while other approaches are focused on promoting creativity among students and preparing them to be problem solvers. One of these approaches that is implemented in classrooms to increase creativity is the Real Engagement in Active Problem Solving (REAPS) model. The REAPS model has been implemented in various counties, including the United States, China, Saudi Arabia, New Zealand, Australia, and Indonesia (Alhusaini, 2016; Alhusaini, Maker, & Alamiri, 2015; Gomez-Arizaga, Bahar, Maker, Zimmerman, & Pease, 2016; Maker, Zimmerman, Alhusaini, & Pease, 2015; Riley, Webber, & Sylva, 2017; Reinoso, 2011; Webber, Riley, Sylva, & Scobie-Jennings, 2018; Wu, Pease, & Maker, 2015; Yulindar, Setiawan, & Liliawati, 2018; Zimmerman, Maker, Gomez-Arizaga, & Pease, 2011). The purpose of this model is to develop students' ability to solve real problems in creative and effective ways in different domains.

During my PhD studies, I worked with my academic advisor on different projects, one of which was implementing the REAPS model in one public elementary school in Australia. I was interested in examining the effects of this model on students' creative problem solving in science and linking the outcomes of this model with the fidelity of implementation (FOI) to ensure that the teachers implemented the model in a way that was described by its developers. Measuring FOI and connecting it with the outcomes of the REAPS model increased confidence that these outcomes came from the effectiveness of this model rather than from other external variables. Educators are now working in an era of research-based evidence, so they must be careful to interpret data based on its context.

I had many purposes in this study. First, I examined the influence of levels of FOI on students' creative problem solving in science. In several studies, a high level of FOI was associated with desirable outcomes while a low FOI was not linked with desirable results. Second, I measured the relationship between creative problem solving in science and grade levels to see if the growth of creativity in science is based on developmental age. Different researchers had arrived at different conclusions about the growth of creativity at certain ages. For example, a researcher found that creativity increased among students in grade 4 while one researcher concluded that creativity in grade 4 declined. Third, I examined whether students' creative problem solving ability in science increased or decreased based on the teachers' levels of FOI during two years of implementing the REAPS model. For example, if students were with a low implementer in Year 1 and then moved to a classroom with a high implementer in Year 2, did their creative problem solving ability in science increase or decrease because of these changes compared with students who were with high implementers in Years 1 and 2? Finally, I investigated which factors of the Test of Creative Problem Solving in Science (including Finding Problems, Generating Detailed Solutions, and Classifying Elements) were most associated with the level of FOI.

In the next part of this chapter, I discussed three areas that included creativity, the REAPS model, and FOI, and I provided background for these areas. In the creativity section, I covered its definitions, how it occurs, the relationship between creativity and developmental age, creative problem solving, and problem types. In the REAPS model section, I introduced this model, its components, and studies that provided support for the positive impact of the model. In the FOI section, I discussed its definitions, its components, factors that affected it, the importance of measuring FOI, and the relationship between levels of implementation and outcomes.

Creativity

Everyone is creative in some way (Isaksen, Dorval, & Treffinger, 2011; Tracy, 2015); however, people's misunderstanding of what creativity is may limit their own creativity. The reasons behind this may be attributed to the following assumptions: (a) only a few individuals are creative, such as gifted people, (b) creativity is associated with particular disciplines, such as art, design, and advertising, (c) creative ideas come to people quickly and without any effort, (d) creativity is associated with genes, so people who are creative will have creative children as well, (e) creative ideas should not be built from other previous ideas, and (f) creative ideas come from experts, so novice individuals cannot be creative (Burkus, 2014; Isaksen et al., 2011; National Advisory Committee on Creative and Cultural Education [NACCCE], 1999).

Definition of Creativity

Creativity has been researched for many decades, and scholars from various disciplines, such as psychology and education, have attempted to develop a comprehensive definition of creativity that is widely recognized among researchers (Amabile, 1983; 1996; 2013; Csikszentmihalyi, 2006; Torrance & Shaughnessy, 1998). However, no consensus definition of creativity has been established (Kerr, 2009; Isaksen et al., 2011). For example, Csikszentmihalyi defined creativity as “a process that can be observed only at the intersection where individuals, domains, and fields interact” (p. 3). Domain refers to all cultural aspects such as physics or art, whereas field refers to social aspects such as teachers, critics, and specialists in domains who make decisions about what to include in a domain and what not to include.

Another attempt to define creativity came from a scholar who spent more than 30 years studying it (Amabile, 1983; 1996; 2013). She defined creativity as “the production of a novel and appropriate response, product, or solution to an open-ended task” (2013, p. 134). Torrance (1988), who developed a well-known test for creativity called the Torrance Tests of Creative Thinking (TTCT), defined creativity as “the process of sensing difficulties, problems, gaps in information, missing elements, something askew; making guesses and formulating hypotheses about these deficiencies, evaluating and testing these guesses and hypotheses; possibly revising and retesting them; and finally communicating the results” (p. 47). Maker (1993) defined giftedness as the ability to solve complex or simple problems “in the most efficient, effective, or economical ways.” (p. 71). Finally, Hu and Adey (2002) proposed that scientific creativity is “a kind of intellectual trait or ability producing or potentially producing a certain product that is original and has social or personal value, designed with a certain purpose in mind, using given information” (p. 392).

Common characteristics of creativity can be found among various definitions: (a) creativity is a series of processes; (b) creativity requires specific characteristics possessed by creative persons, such as imagination and openness; (c) creative outcomes are novel, appropriate, and useful to the creator, in particular, and to the audience, in general; and (d) creativity requires interaction between individuals and their environments to generate unique outcomes (Amabile, 1996; Guilford, 1950; Isaksen et al, 2011; Mednick, 1962; Rhodes,1961; Torrance,1993; Welsch, 1980; Zeng, Proctor, & Salvendy, 2009). Thus, I propose that creative problem solving in science can be defined as the ability to solve a simple or complex scientific problem in the most efficient, effective, or economical way, and these creative outcomes must be original to the relevant developmental stage of the individual or audience.

How Does Creativity Occur?

Many scholars have proposed models and theories to illustrate and explain the processes that lead to the developing of creative ideas and products (Amabile, 1983; 1996; 2013; Csikszentmihalyi, 2006; 2014). For example, Csikszentmihalyi suggested a model of creativity that he called a "systems model of creativity" (Figure 1).

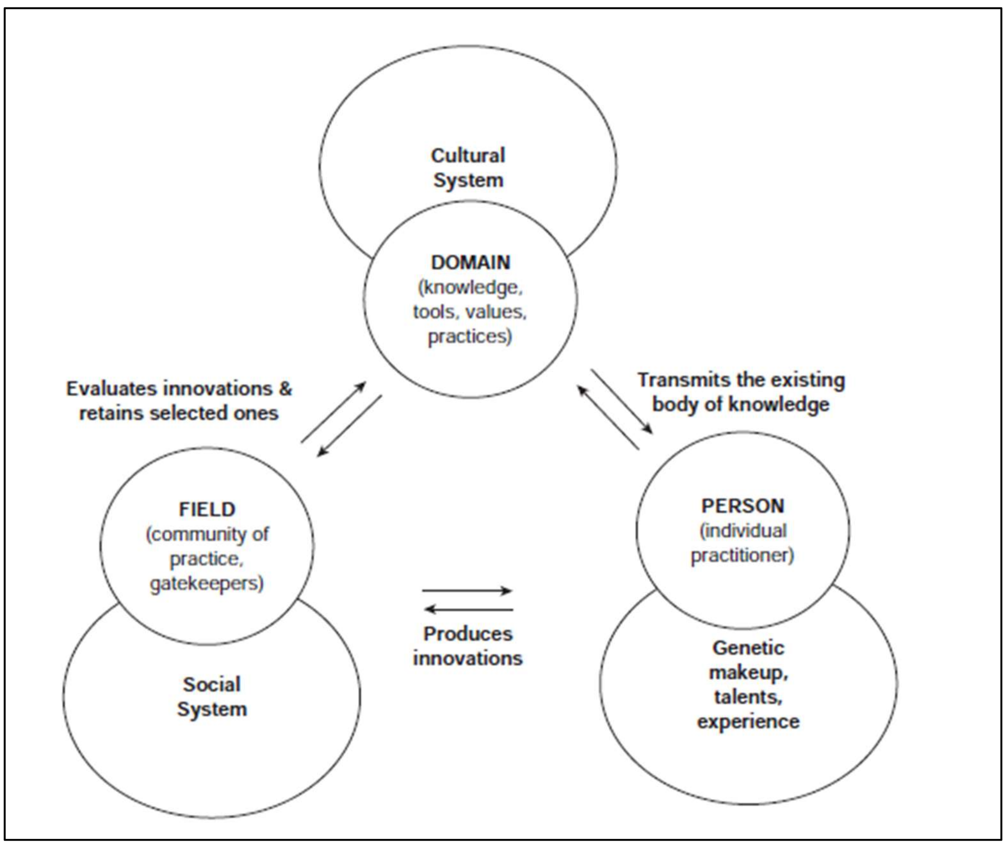

Figure 1. The Systems Model of Creativity

Note: The figure was introduced by Csikszentmihalyi, M. (2006) A systems perspective on creativity. In: Henry J (3rd ed) *Creative management and development*. Sage Publications, London: UK, pp 3–17

In his framework, Csikszentmihalyi (2006; 2014) identified three elements associated with creativity: *domain, person*, and *field*. Domain refers to any branch of knowledge such as science, art, and mathematics. Person refers to a creative individual who produces creative ideas or products. The person may be affected by different factors such as family genetics and experiences from his domain. Field refers to experts in specific domains who make decisions about creative ideas or products, such as scientists, teachers, or artists. These elements must interact with each other to produce creativity. Each element plays an important role in creativity but not enough to produce originality itself (Csikszentmihalyi).

In Csikszentmihalyi's model, creativity occurs when an individual obtains insight or wisdom from his or her practice within the domain and then produces "a novel variation in the content of the domain." The variation then must be selected by the field for inclusion in the domain (Csikszentmihalyi, 2006, p. 3). For example, when an individual creates a new idea (person) in physics (domain), this idea is then reviewed by physics scientists (field), who decide whether to accept it; thereafter physics scientists in general adopt this idea in their practice. This process can be seen, for example, in the acceptance of Einstein's theory of relativity: first, Einstein had an idea, then a small number of physics professors approved his novel idea as a creative contribution to their field of physics, finally millions of people accepted this judgment and adopted the idea into their work (Csikszentmihalyi, 2006).

Another example is the Componential Theory of Creativity (Figure 2) which was developed by Amabile (1983; 1996; 2013). This model's definition of creativity was "the production of ideas or outcomes that are both novel and appropriate to some goal" (Amabile, 2013, p. 134). In her model, creativity has four components, three of which belong to creative individuals: *Domain-Relevant Skills, Creativity-Relevant Processes,* and *Task Motivation*. The

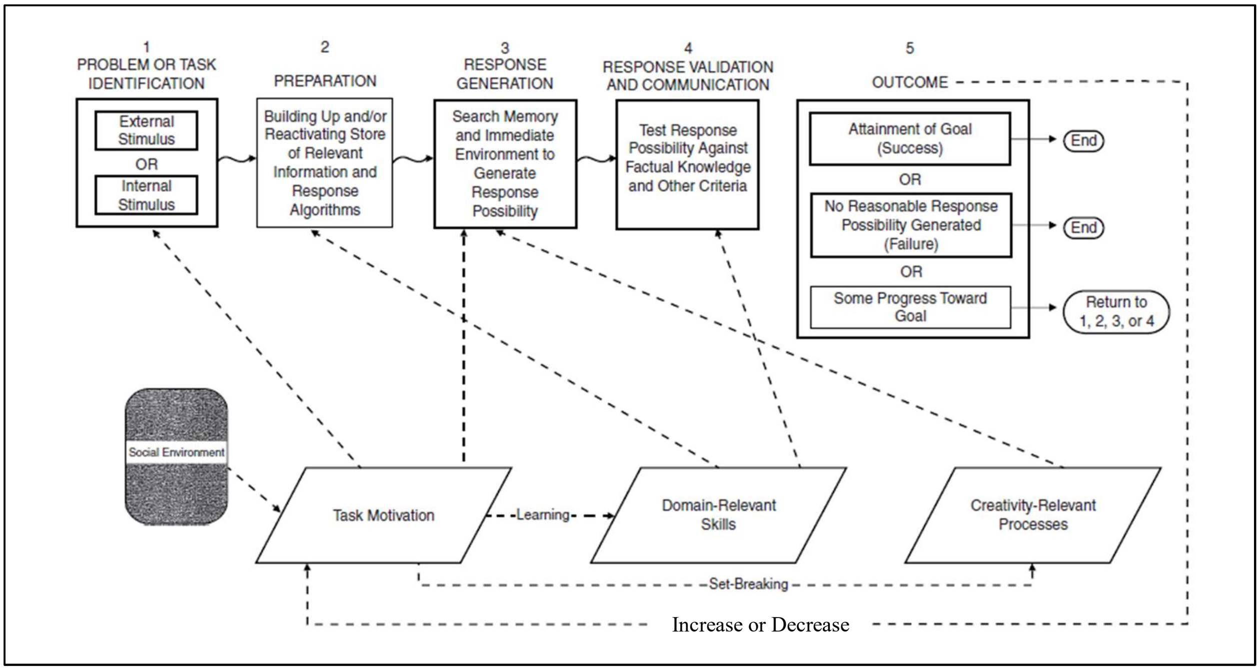

*Figure 2***.** The Componential Theory of Creativity

Note: The figure was introduced by Amabile, T. M. (2013). Componential theory of creativity. In *Encyclopedia of management theory.* (pp. 134-139). Los Angeles, CA: SAGE

fourth component is *Social Environment*, the environment in which the individual works. Domain-Relevant Skills are the fundamental components of any performance in any domain. They include many factors that affect the domain, such as a person's talent and intelligence, the knowledge this person gains regarding the particular domain, the personal expertise developed over the years, and the technical skills that are essential for the domain such as laboratory techniques.

Creativity-Relevant Processes include two factors: *cognitive style* and *personal characteristics of creative problem solvers*. These two factors necessitate that a creative individual express key characteristics such as "independence, risk-taking, and taking new perspectives on problems, as well as disciplined work style and skills in generating ideas" (Amabile, 2013, p. 135). Cognitive style includes the ability to "use wide, flexible categories for synthesizing information and the ability to break out of perceptual and performance scripts" (p. 135). The personal characteristics of creative problem solvers include tolerance for ambiguity and self-discipline.

Task Motivation is the spark that leads to the production of creative ideas or outcomes. It includes a set of internal factors that encourage creative individuals to complete the tasks that result in creative outcomes. In Task Motivation, "[p]eople are most creative when they feel motivated primarily by the interest, enjoyment, satisfaction, and challenge of the work itself" (Amabile, 2013, p.136).

Social Environment plays a significant role in the school setting or other workplace. It includes external motivations such as encouragement from teachers, parents, and company managers. However, some factors may decrease creativity, such as criticizing new ideas harshly, not encouraging risk taking, and not allowing sufficient time.

Creativity and Developmental Age

An association has been found between the growth of creativity and ages from childhood to adolescence, with some exceptions at certain developmental ages (Maker, Jo, & Muammar, 2008; Urban, 1991; Torrance, 1963; 1968). In several studies, the association between age and creativity among students in elementary schools was mixed. For example, a study was conducted to assess the effectiveness of a curriculum teaching model, Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses (DISCOVER), on students' creativity from kindergarten to grade six in four elementary schools in the United States. The researchers used the Test of Creative Thinking-Drawing Production (TCT-DP) to measure students' general creativity after they had been exposed to the teaching model for three years. The researchers found, in general, that students' creativity increased from year to year with no "significant peaks or slumps" except for students at age six where their creativity increased significantly (Maker et al., 2008, p.414). The results of this study were different from those of another study of general creativity using the TCT-DP with children ages 4 to 8. The researcher observed a direct correlation between creativity and student age; however, creativity declined at age 6 (Urban, 1991). Torrance (1968) also found in his longitudinal study that creativity among students in grade 4 decreased compared with students in grades 3 and 5.

The decline of creativity at certain developmental stages may be attributed to a number of factors. First, formal schooling may be a significant factor that affects creativity among children (Maker et al., 2008). For example, a researcher compared results from two groups of children at age 6, finding that children who were in kindergarten had a higher mean creativity ($M = 28.2$) than children at age 6 in first grade ($M = 14.5$; Urban, 1991). Second, in literature on creativity, "authorities and circumstances of socialization" play a significant role in decreasing creativity among children because accepting creative ideas and products is mostly "dependent on environmental conditions, including educational institutions" (Urban, p. 178). Finally, the structure of school may play a role in increasing or decreasing creativity among students (Thomas & Berk, 1981). For instance, a researcher compared students' creativity in two classrooms, an open classroom and a traditional classroom, defining open classroom as a class with "a diversity of materials and flexibility of procedures designed to support each child's individual abilities" (Hyman, 1978, p. 268). Students' creativity in the open classroom was higher than that of students in the traditional classroom (Hyman). However, these results were contradicted by those of another study, in which student creativity in the traditional classroom was higher than in the open classroom (Wilson, Stuckey, & Langevin, 1972).

Creative Problem Solving

Everyone faces problems in daily life, both simple and complex. Thus, acquiring problem solving skills is essential to everyone. But first, what is a problem? In general, it has several meanings based on the context. Its synonyms include difficulty, question, trouble, predicament, dilemma, imbroglio, mess, and complication (Urdang, The Oxford Thesaurus, 1993). One of the most widespread meanings of problem is “a question raised for inquiry, consideration, or solution” (merriam-webster.com, 2018). Problem does not necessarily connote something negative. In some cases, problem means simply any challenge that needs to be resolved (Maker, 2006). For example, in a workplace, one may suggest a modification to a process to increase the quality of production even though the current system works well.

Another question arises. What does problem solving mean? Many experts in the fields of education and psychology, for example, have introduced several definitions and explanations of problem solving (Gardner, 1983; Maker, 2006; Sternberg, 1985; 2005). Maker defined problem solving as “the process of answering questions, resolving difficulties, creating solutions, and investigating perplexing situations” (Maker, 2006, p. 38). Gardner, who developed the well-known Theory of Multiple Intelligences, described intelligence as the ability to solve problems. Sternberg also described problem solving as a set of components that includes metacomponents, performance components, and knowledge-acquisition.

To solve any problem, individuals tend to use different thinking strategies to reach the most appropriate and effective solutions. In general, two approaches can be employed to solve any problem or challenge: *convergent* and *divergent thinking* (Bachelor & Michael, 1991; Guilford, 1959; 1981; 1982; 1988; Mumford & Gustafson, 1988).

Convergent thinking. Convergent thinking is a process of finding a single, most correct and appropriate solution to a given question or problem (Bachelor & Michael, 1991; Cropley, 2006; Guilford, 1959; 1981; 1982; 1988; Mumford & Gustafson, 1988). Individuals who embrace convergent thinking use a combination of logic, speed, and accuracy to reach solutions to their problems (Cropley). This approach is appropriate when there is a single, previously-known answer that "needs simply to be recalled from stored information or worked out from what is already known by applying conventional and logical search, recognition, and decision-making strategies" (Cropley, p. 391). A good example of convergent thinking can be found on standardized multiple-choice tests in which examinees select only one of the proposed answers for each question.

Divergent thinking. Divergent thinking is a process of generating many solutions or ideas for a given problem or situation (Bachelor & Michael, 1991; Guilford, 1959; 1981; 1982;1988; Mumford & Gustafson, 1988; Saccardi, 2014; Scratchley & Hakstian, 2000). An example of divergent thinking can be found during an open-ended test in which individuals generate as many possible solutions as they can to solve a challenging situation. Divergent thinking is the most vital strategy people use to create novel ideas or products (Baer; 1993, Cropley, 2006). In fact, it is more associated with creativity than is convergent thinking (Baer; Cropley; Vincent, Decker, & Mumford, 2002), especially when creativity means novel and unusual ideas or products (Gibson, Folley, & Park, 2009).

Components of divergent thinking. Divergent thinking includes four components: *fluency*, *flexibility*, *elaboration*, and *originality* (Bachelor & Michael, 1991; Guilford, 1959; Mumford & Gustafson, 1988).

Fluency. This is the ability to generate or produce many problems or solutions to a given question or situation (Bachelor & Michael, 1991; Guilford, 1959; Mumford & Gustafson, 1988). Fluency is considered the first step in solving a problem and producing a creative idea or product (Shively, 2011). One example is a classroom teacher showing a video about the global revolution of the smartphone and then asking students to identify as many problems as possible about the use of smartphones.

*Flexibility***.** This is the ability to sort many ideas into categories based on the similarities of the ideas (Bachelor & Michael, 1991; Guilford, 1959; Mumford & Gustafson, 1988; Shively, 2011).

Elaboration. This is the ability to add a meaningful description to an idea or to complete a creative idea (Bachelor & Michael, 1991; Guilford, 1959; Shively, 2011). Elaboration is an effective strategy for illustrating and expanding an idea to make it clearly understandable for other individuals (Shively).

Originality. This is the ability to create or produce a new and novel idea or product (Bachelor & Michael, 1991; Guilford, 1959; Mumford & Gustafson, 1988). Originality is the component most associated with creativity (Gibson et al., 2009). How do researchers determine whether an idea or product is original? In general, they need to set the criteria for the ensuing discussion about the originality of an idea or product (Hu & Adey, 2002; Maker, Jo, Alfaiz, & Alhusaini, 2017). One way to set the criteria for determining the originality of a set of ideas is to use a statistical approach to distinguish between original and common ideas. Some researchers sort all ideas based on their frequencies among participants and then select ideas from the top five percent, assigning points to the novel ideas that were located within this top five percent (Salemi, 2010).

Another way of assessing originality is by using the Consensual Assessment Technique (CAT; Amabile, 1983; Lubart, 1994). This strategy is appropriate for rating students' products in classrooms. In this strategy, two or more judges (raters) are asked to give their assessments of creative products using a 7-point scale, with one point meaning that the product is less creative and seven points indicating the most creative of the products. However, the judges must select at least one product for each point.

Amabile (1983) suggests requirements for judges who make decisions about creative products: (a) they should be external observers "who have not been preselected on any dimension other than their familiarity with the domain." (p. 38); (b) they must make their decisions independently; (c) they should rate other dimensions of creative products in addition to creativity of the products (e.g., if students develop products by writing on papers, the judges can rate the quality of students' handwriting or their use of fundamentals such as grammar and punctuation). The purpose of asking the judges to rate other dimensions is to ensure they are not influenced by the other factors; (d) they should be "instructed to rate the products relative to one another on the dimensions in question, rather than rating them against some absolute standards they might have for work in their domain." (p. 38); and (e) they should rate creative products in random order.

Problem Types

Problem types are associated with the convergent and divergent thinking approaches. In general, there are two main types of problems: closed and open-ended. Convergent thinking is mostly connected to closed problems because solvers seek to find the most correct solution. On the other hand, divergent thinking is more relevant to open-ended problems because solvers produce many possible answers and then select one or more. Researchers have proposed strategies to present a problem to its solvers (Getzels & Csikszentmihalyi, 1967; 1976; Maker & Schiever, 2005; 2010).

Getzels and Csikszentmihalyi (1967; 1976) introduced three types of problems ranging from closed to open-ended. In the first type, both the problem presenter and the solver know the problem and the method of solving it. However, only the presenter knows the correct answer in advance, while the solver needs to figure out the correct solution. In the second type of problem, the presenter and solver know the problem but only the presenter knows the method and solution. Thus, the role of the solver is to determine the method and solution. In the third type, the role of the solver is to identify a problem and propose a method for solving it.

Maker and Schiever (2005; 2010) suggested a new way to expand the work of Getzels and Csikszentmihalyi (1967; 1976). They developed six problem types (Table 1). The first two types are closed problems. Type I is similar to Type I in Getzels's and Csikszentmihalyi's work in which only one answer can be identified by the solver. Types III and IV are semi-open problems. Types V and VI are open-ended problems. The reason that Maker and her colleagues grouped the six types into three levels (closed, semi-open, and open-ended problems) was to make it easier for teachers to develop classroom activities appropriate for the different problem types. The purpose of applying closed problems in classrooms is to assist teachers with assessing and developing student knowledge and skills, while the purpose of designing activities based on open-ended problems is to assess and develop students' creative problem solving abilities (Maker, 2017a; 2017b).

Table 1: *Problem Types*

Problem	Type	Problem		Method		Solution	
		Presenter	Solver	Presenter	Solver	Presenter	Solver
Closed	I	Specified	Known	Known	Known	Known	Unknown
	II	Specified	Known	Known	Unknown	Known	Unknown
Semi-open	III	Specified	Known	Range	Unknown	Known	Unknown
	IV	Specified	Known	Range	Unknown	Range	Unknown
Open-ended	V	Specified	Known	Unknown	Unknown	Unknown	Unknown
	VI	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown

Note. Adapted from "Problem Continuum" by Maker, J., & Schiever, W. (2010). *Curriculum development and teaching strategies for gifted learners* (3rd ed.). Austin, TX: Pro Ed.

The REAPS Model and Evidence of Its Effectiveness

The Real Engagement in Active Problem Solving (REAPS) model was developed by Maker and Zimmerman in 2004 to improve students' abilities to solve real-world problems in effective ways. This model can be implemented in any educational environment from elementary to high school, in different cultural contexts, and with a variety of educational programs (Maker & Zimmerman, 2008, Maker et al., 2015). In different studies, students expressed their belief that the REAPS model was a means to help them engage in the learning process, challenge their abilities, and develop their interpersonal skills (Gomez-Arizaga et al., 2016; Wu et al., 2015). Similarly, teachers found the REAPS model to be an effective teaching model that enhanced their teaching (Riley et al., 2017; Zimmerman et al., 2011).

The REAPS model is a combination of three teaching models: Problem-Based Learning (PBL), Thinking Actively in a Social Context (TASC), and the Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses (DISCOVER). The three models of teaching complement each other, and each one of them has its contribution in the REAPS model (Maker & Zimmerman, 2008).

Problem-Based Learning (PBL)

The purpose of PBL is to provide opportunities for students to solve complex real-world problems (Maker & Zimmerman, 2008). Teachers who use PBL need to select an undefined problem that relates to a real life situation. The role of the teachers is to facilitate the learning of their students, while the students work in small groups and play the role of stakeholders, integrating their specific knowledge from different disciplines to solve the problem (Maker et al., 2015). Students can investigate the problem further in different ways, such as with field trips, discussions, and lectures, to better their understanding (Maker & Zimmerman).

The effectiveness of the PBL model was supported in different studies. The positive impact of this model on students' motivation, self-regulation, and argumentation abilities was reported by several researchers (Belland, Glazewski, & Richardson, 2011; Ersoy & Baser, 2010; Gallagher & Gallagher, 2013; Jo & Ku, 2011).

Thinking Actively in a Social Context (TASC)

The main goal of TASC is to develop students' abilities to solve real-world problems (Maker & Zimmerman, 2008). The TASC model contains eight steps: Gather/Organize, Identify, Generate, Decide, Implement, Evaluate, Communicate, and Learn from Experience (Adams & Wallace, 1991; Wallace, Bernardelli, Molyneux, & Farrell, 2012). Students can alternate between these steps to further investigate and solve problems (Maker et al., 2015). For example, students can gather and organize information about a certain problem several times in different steps to ensure that their current solution is an appropriate one.

Evidence of the effectiveness of the TASC model was found in classrooms. Overall student achievement improved (West, 2008), and students expressed positive attitudes toward learning, especially students from low socioeconomic levels (Fitton & Gilderdale, 2008). Students in schools where teachers implemented the TASC model also became more independent learners, were more engaged in their own learning, and enjoyed working with their peers in groups (West). Students believed that the TASC model was instrumental in helping them learn how to solve problems and apply thinking processes to create solutions (Ball & Henderson, 2008).

Likewise, after implementing it in their classrooms, teachers expressed positive attitudes toward the TASC model (Willmoth, 2008) and that it was easy to implement (Ball & Henderson, 2008). The teachers believed their commitment and dedication to teach students increased, and this model was instrumental in helping them work with other teachers to design lessons and combine different academic subjects (Cartwright, 2008; Fitton & Gilderdale, 2008).

The DISCOVER Curriculum Model

The DISCOVER curriculum model was developed by Maker and her colleagues (1994) to enhance the quality of teaching gifted students from different ethnicities, languages, and backgrounds. The DISCOVER curriculum model is based on the following ideas: (a) the Problem Solving Continuum of Getzels and Csikszentmihalyi (1967), (b) the Multiple Intelligences Theory (Gardner, 1983), (c) the Triarchic Theory of Intelligence (Sternberg, 1985), and (d) specific research in education of the gifted, as well as research on creativity and its development (Maker et al., 2006). The use of the DISCOVER curriculum model can foster a positive attitudes toward learning more than other teaching models, in several ways: (a) providing opportunities for students to develop diverse abilities and interests, (b) giving students chances to solve real world problems, (c) integrating hands-on learning activities into lessons, (d) incorporating students' cultural awareness into the curricula, and (e) designing curricula around state standards and abstract themes (Maker et al., 2015).

Teachers can design their lessons by using the problem types (Maker & Schiever 2005; 2010). Several researchers supported implementation of DISCOVER in classrooms and showed its impact on participants' creativity (Kuang, 2011; Kuo, Maker, Su, & Hu, 2010; Maker et al., 2006; Maker et al., 2008). For example, Maker and colleagues (2006) found that students whose teachers implemented the DISCOVER curriculum model at a high or middle level showed increases in their levels of creativity compared to students whose teachers implemented the DISCOVER curriculum model at a low level.

Fidelity of Implementation

How do we know whether an education program or teaching model achieves its objectives? Do we need to acquire scientific evidence of the effectiveness of a teaching model? Only with evidence of fidelity of implementation (FOI) of any program can policymakers, educators, and parents be confident that the outcomes of the program are due to its design rather than to other external variables.

Fidelity of implementation first appeared in the 1970s in the behavioral intervention literature (Bellg et al. 2004). However, only in 1991 were Moncher and Prinz (1991) the first authors to define this concept and suggest guidelines for improving it. Before the 1970s, researchers expected that implementers of a certain program or intervention would copy and replicate it with high FOI as described by its developers (Rogers, 1983). This expectation existed because implementers were "considered to be rather passive acceptors of an innovation, rather than active modifiers of a new idea" (Rogers, p. 176).

The assumption of FOI is that a high level of implementation will lead to high or desirable outcomes while a low level of implementation will produce poor or undesirable results (Harn, Parisi, & Stoolmiller, 2013; Maker et al., 2006; Maker et al., 2008). This assumption created an overall agreement between researchers and practitioners "on what fidelity is, how to measure it, and what level of fidelity optimizes outcomes." However, no widespread agreement on any of these definitions was established (Harn et al., p. 181). In fact, when the components of a program are defined well by its developers, the opportunity for implementing the program successfully is increased. In other words, implementers' knowledge of the core components of a specific program or practice will increase the chance for successful implementation (Fixsen, Naoom, Blase, & Friedman, 2005).

The Definitions of Fidelity of Implementation

What is FOI? No singular definition of FOI has been established (Dusenbury, Brannigan, Falco, & Hansen, 2003). As a result, several different definitions of FOI exist among various disciplines such as health and psychology. For example, Przybylski and Orchowsky (2015) define FOI as "the degree to which a program's implementation in any real-world setting matches what was stated in the original program model" (p. 3). Another definition was suggested by Fixsen et al. (2005), who state that FOI is "a specified set of activities designed to put into practice an activity or program of known dimensions" (p. 5).

In the field of education, a number of definitions of FOI have emerged (O'Donnell, 2008). For instance, the National Research Council (NRC; 2004) defined FOI as "a measure of the basic extent of use of the curricular materials. It [FOI] does not address issues of instructional quality. In some studies, implementation fidelity is synonymous with 'opportunity to learn'" (p. 114). Another definition was proposed by Fullan (2001), who defined FOI as a method "to get individuals and groups of individuals to implement [the program] faithfully in practice – that is, to use it as it is [supposed to be used] as intended by the developer" (p. 25). In short, based on the various definitions of FOI that have been discussed, we can conclude that FOI is the degree to which implementers such as teachers apply a program or teaching model in the manner that is described by the program developers (Dumas, Lynch, Laughlin, Smith, & Prinz, 2001; Dusenbury et al., 2003; Mowbray, Holter, Teague, & Bybee, 2003).

The Importance of Assessing Fidelity of Implementation in Research

Assessing FOI is important for a number of reasons (Harn et al., 2013; NRC, 2004; Nelson, Cordray, Hulleman, Darrow, & Sommer, 2012; No Child Left Behind Act, 2001; O'Donnell, 2008). First, assessment of FOI helps researchers understand how the critical components of their programs are carried out and how these components can be improved, especially when the researchers plan to expand and disseminate the program nationally or internationally. One benefit of assessing FOI is that researchers and policymakers can expand the implementation programs to new settings, especially if these programs produce desirable outcomes and are implemented with high fidelity (Swanson, Wanzek, Haring, Ciullo, & McCulley, 2013).

Second, such evidence is critical in evaluating whether the observed program outcomes are the result of the program interventions themselves or of other external variables (Harn et al., 2013; O'Donnell, 2008). Fidelity of implementation gives researchers opportunities to determine what has been changed in their programs by the instructors and how these changes influence outcomes (Harn et al.). Assessing FOI helps researchers to understand how the quality of implementation affects outcomes. Teachers who share the philosophy of the intervention are more likely to implement it with a high level of fidelity than teachers who do not share the philosophy (Harn et al.). In fact, assessing FOI provides valuable information that can help describe why results do or do not occur (Duerden & Witt, 2012). Researchers pointed out that one program may produce different results because of the site of implementation (Durlak, 1998).

Third, FOI assessment is a response to the call for accountability in K-12 curriculum interventions research (NRC, 2004). Many policymakers and experts have been asking researchers to provide scientific evidence of fidelity of their interventions and the efficacy and effectiveness of the outcomes. Teachers are asked in government initiatives such as the No Child Left Behind Act (2001) to use and implement only those research-based teaching methods that have supportive evidence of their effectiveness. In this era of research-based evidence, providing scientific evidence of the effectiveness of a teaching program is critical for accepting its results (Mowbray et al., 2003). Fidelity of implementation is considered a core element for implementing evidence-based practices (Blandford & Osher, 2012; Przybylski & Orchowsky, 2015; Welsh, Sullivan, & Olds, 2010). The lack of evidence of FOI is associated with poor outcomes in evidence-based practices studies (Welsch et al.).

Finally, assessing FOI provides valuable evidence of the feasibility of a program. When researchers fail to achieve FOI in practice, their programs may have low levels of feasibility (Nelson et al., 2012). Assessing FOI allows researchers to offer clear explanations for why the program succeeds or fails (Nelson et al.). When a program that is implemented with a high level of fidelity fails to produce the anticipated outcomes, researchers may need to redesign it (Nelson et al.).

Components of Fidelity of Implementation

Any teaching model has core components, and teachers attempt to implement the model with high fidelity to ensure that all students benefit from the model and improve their academic performance. However, students' abilities and needs vary, so teachers try to adjust the teaching model to fit the needs of their students. The challenge that teachers face is finding ways to balance a strict implementation of the teaching model with making some adjustments to meet students' needs (McMaster et al., 2014). In this case, the developers of the teaching model need to distinguish the core components of their model that teachers must implement with fidelity from the other elements of the model with which teachers have the flexibility to adjust as needed (McMaster et al.).

To increase chances for a high FOI, researchers may consider a proposal that was developed by Fixsen, Blase, Naoom, and Wallace (2009) from the first stage of planning their program until implementing it. The proposal contains the following components: *Staff Selection*, *Preservice and In-Service Training*, *Ongoing Coaching and Consultation*, *Staff Performance Assessment*, *Decision Support Data Systems*, *Facilitative Administration*, and *Systems Intervention* (Figure 3).

Staff Selection. Researchers need to specify the criteria for selecting the implementers of their interventions or programs. The following are examples of questions that help to set the selection criteria: (a) "Who is qualified to carry out the evidence-based practice or program?" (b) What are the methods for recruiting and selecting teachers with those characteristics? (c) What is the minimum knowledge that teachers should have for optimal implementation? and, (d) What are the basic professional skills that teachers should acquire before implementing the program or intervention? (Fixsen et al., 2009, p. 533).

Preservice and In-Service Training. This component is important to give teachers the necessary background knowledge of the theory and philosophy behind the program or intervention they are going to implement in their classrooms. During the training sessions,

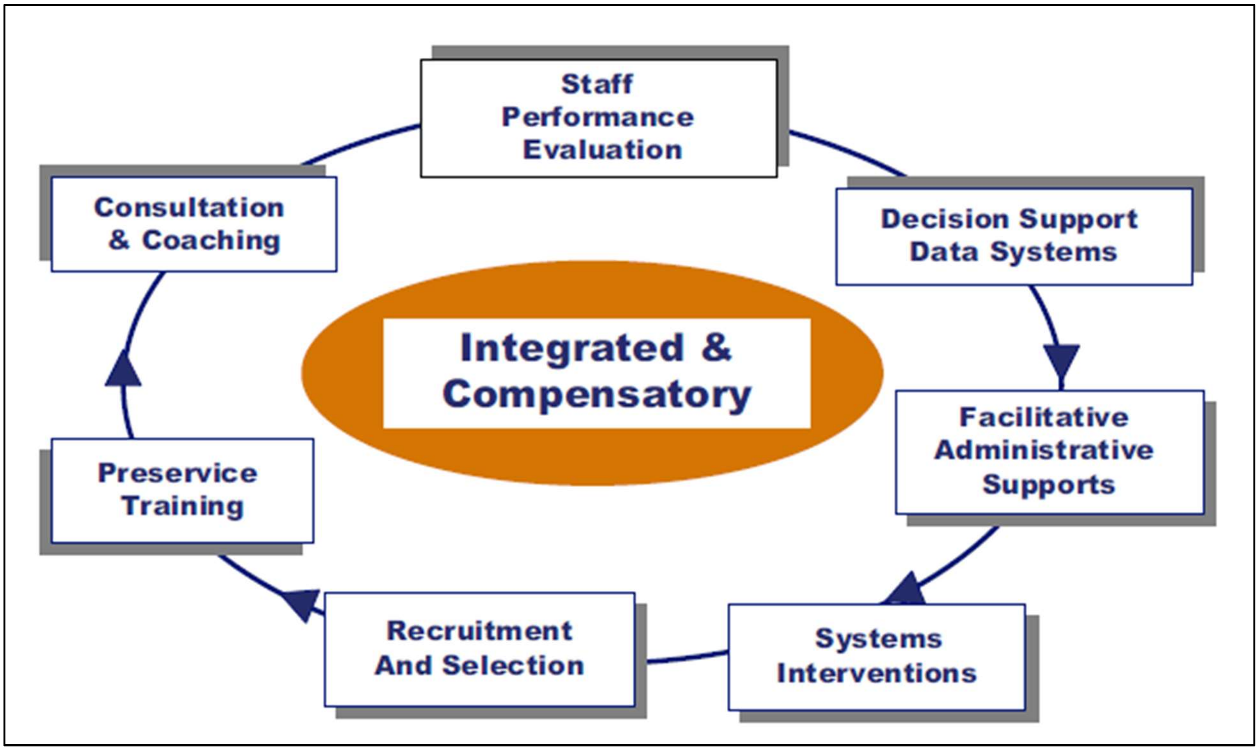

Figure 3. Core Implementation Components

Note. The figure was introduced by Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on Social Work Practice, 19*(5), 531-540. 10.1177/1049731509335549

teachers will know how to implement the program and have a chance to practice the skills needed for best implementation. This training provides a great opportunity for the teachers to receive feedback from the researchers (developers) of the program and give them a chance to practice the necessary skills before applying them in the classroom.

Ongoing Coaching and Consultation. Although providing training sessions to the teachers prior to implementation is important, some researchers emphasized the role of coaching and consultation after training (Joyce & Showers, 2002; Schoenwald, Sheidow, & Letourneau, 2004; Sholomskas et al., 2005). Coaching and consulting benefit teachers in their implementation of the intervention.

Staff Performance Assessment. This component helps teachers to improve their performance in their classrooms. When researchers assess teachers in the classroom and measure FOI, teachers receive feedback from the researchers regarding possible ways to improve their practices to meet the program's standards.

Decision Support Data Systems. The developers of a program need to develop a strategy to measure "key aspects of the overall performance of the organization and provide data to support decision-making to assure continued implementation of the core intervention components over time" (Fixsen et al., 2009, p. 535).

Facilitative Administration. Teachers need to receive supports from school principals and other school staff to implement the program in an effective way. The school principals can "support the overall processes, and keep staff organized and focused on the desired intervention outcomes" and help teachers to solve any challenges they may encounter (Fixsen et al., 2009, p. 535).

Systems Intervention. This includes any strategy to deal with any external systems to "ensure the availability of the financial, organizational and human resources required" to assist teachers in the best implementation of the program in their classrooms (Fixsen et al., 2009, p. 535).

Factors That Affect Implementation

Three factors play significant roles in increasing the chances for better implementation of any program with high fidelity: *Community Level Factors*, *Provider Characteristics*, and *Innovation Characteristics* (Durlak & DuPre, 2008). Community Level contains four sub-factors, which include the prevention research system, politics, funding, and policy factors. The factors associated with Provider Characteristics are: (a) perceived need for innovation and its relevance to local needs, (b) perceived benefits of innovation and whether “the innovation will achieve benefits desired at the local level” (Durlak & DuPre, p. 337), (c) self-efficacy and whether teachers are willing to implement program components as described by the developers, and (d) skill proficiency and whether the teachers possess skills necessary for implementation. The two Innovation Characteristics are: (a) compatibility and whether the intervention reflects the mission, value, and priorities of the organization and (b) adaptability of the program to be modified and fit with “provider preferences, organizational practices, and community needs” (Durlak & DuPre, p. 337).

Measuring Fidelity of Implementation

One important purpose of measuring FOI is to acquire evidence of the internal validity of a study and increase the confidence that the study outcomes relate to its intervention variables and not to extraneous variables (Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000). For instance, if a desirable outcomes ensue from a study but the intervention of the study was not implemented as proposed by its developers, then possible reasons for the significance of the results may be “maturation, instruction in the general education environment, a substantively altered version of the intervention, or any other element of schooling” (Harn, et al., 2013, p. 182).

In several studies, categorical and continuous techniques were the most widely used approaches to measure FOI. In the categorical technique, researchers divided implementers into two or more groups such as high and low based on implementation differences. In the continuous technique, researchers evaluated teachers based on their percent of total fidelity of implementation (Durlak & DuPre, 2008). Most studies used a categorical technique to measure

FOI. For example, in over 59 studies that Durlak and DuPre reviewed, 58 percent of the authors used a categorical technique (e.g., low versus high implementation groups).

The method of measuring FOI is one of the most important aspects of implementation research. While multiple forms of data collection have been developed, the most popular methods for measuring FOI are ratings by experts, interviews, and self-reports by implementers (Durlak & DuPre, 2008; Dusenbury et al., 2003; Mowbray et al., 2003).

Ratings by experts can take different forms such as classroom observations, reviewing documentations and program materials, and videotaped sessions. Observation is considered the most valuable and reliable method for assessing FOI (Duerden & Witt, 2012). However, it is the most expensive method for collecting data (Duerden & Witt). One reason for using observations to determine FOI is that this method is more objective than self-reporting instruments (Durlak & DuPre, 2008). To benefit from observational data, researchers are advised to link the data with program outcomes (Durlak, 1998). The most common concern about observations is how often the researchers need to visit classes and observe teachers to measure FOI (Mowbray et al., 2003). In fact, no one single answer is appropriate. The number of observations needed to measure FOI depends on many factors such as student characteristics, school schedules, and teacher availability (Harn et al., 2013; Zvoch, 2009).

Interviewing is considered an appropriate tool to gather in-depth information from participants about the objectives of research (Seidman, 2013). It helps researchers to get answers from teachers about factors that the researchers noticed while they observed classrooms. Interviews can be of different types such as structured and unstructured (Steinbright Career Development Center, 2018). The difference between the two types is that researchers in structured interviews use formal language with the participants and come prepared with open-ended questions (Seidman). Interviews have many advantages; they can clarify any ambiguities that might arise from observing classrooms, allow researchers to achieve a high response rate, and provide rich information. On the other hand, interviews have some limitations such as being time-consuming and costly (Becker & Geer, 1957; Seidman).

Self-report instruments are considered an easy method of acquiring data on implementation, but teachers may misunderstand some items of these instruments or forget some activities they implemented during interventions (Patterson & Graham, 2018). However, they can provide rich information about a program (Nelson et al., 2012). Researchers must be conscious of limitations of self-report instruments: (a) concerns about their reliability and validity, because they tends to skew in a positive direction, (b) the credibility of participants' responses, and (c) variation of interpretations among participants (Durlak, 1998; Fan et al., 2006; Hansen, Graham, Wolkenstein & Rohrbach, 1991; Kaufman, Russell, & Plucker, 2013; Miller, 2014; Silvia, Wigert, Reiter-Palmon, & Kaufman, 2012).

Although attention to assessing FOI has increased significantly in the last few decades, few researchers have proposed valid instruments for measuring it (Mowbray et al., 2003). The challenges for developing a valid and comprehensive measure for assessing FOI may be attributed to two factors: (a) the absence of an established, singular definition of FOI that is accepted by researchers and (b) the varied approaches that researchers use in their studies (Dusenbury et al., 2003).

Until recently, few researchers described any techniques to measure the fidelity of implementation (O'Donnell, 2008). Because FOI is multifaceted and a clear and comprehensive approach to measuring it across various fields such as education and psychology is not available, some confusion exists regarding the best approach for measuring FOI, which results in the reduction of the value and usefulness of fidelity data (Nelson et al., 2012). However, several researchers (Dane, 1997; Dane & Schneider, 1998; Dusenbury et al., 2003; Moncher & Prinz, 1991; O'Donnell) indicated the most common components of studying FOI: (a) adherence, or the extent to which a teaching unit is delivered as designed; (b) duration, or the teacher's use of time management and to what extent this use meets the intent of the developer of the plan, such as the time allocated to deliver concepts or the length of the lessons; (c) quality of delivery, or the way the teacher "delivers the program using techniques, processes, or methods prescribed" (O'Donnell, p. 34) ; (d) participant responsiveness, or the degree to which the students participate in the program's activities; and (e) program differentiation, or "whether critical features that distinguish the program from the comparison condition are present or absent during implementation" (O'Donnell, p. 34).

In some cases, not all of these types must be included to measure FOI and to achieve program outcomes (Dusenbury et al., 2003). Researchers of a single study need to decide which of these five types need to be considered, based on the nature of the study (Dusenbury et al.). The five types can be divided into structure and process (Mowbray et al., 2003). Fidelity of structure contains adherence and duration; fidelity of process includes quality of delivery and program differentiation. The fifth type, participant responsiveness, can be found in any of the two groups (O'Donnell, 2008).

When researchers intend to assess FOI in their programs, they should pay attention to any variables that play critical roles. Some of these variables include teaching experience, quality of instructional leadership, and stability of the setting in which the program is implemented such as principal or teacher change and budget cuts (Harn et al., 2013).

Levels of Implementation and Outcomes

Researchers should be aware of the importance of measuring FOI at any level of their research, from initial pilot to real implementation in actual settings (Durlak, 1998). The programs as implemented in actual settings sometimes result in outcomes different from what was originally intended when the programs were designed. In this case, measuring FOI to increase the confidence in the results of these programs is important. When researchers ignore assessing FOI, they may make incorrect conclusions about the effectiveness of these programs (Nelson et al., 2012). In fact, many funding agencies such as the Institute of Education Sciences (2011) encouraged researchers to measure the efficacy of interventions and "to gather data to help explain the level of fidelity of implementation that is attained to help researchers identify the conditions, tools and procedures that are needed to support implementation of the intervention" (Swanson et al., 2013, p. 3).

In general, a high level of implementation is associated with positive or desirable outcomes. In many studies, researchers have found strong relationships between the level of implementation and outcomes (Durlak, 1998; Durlak & DuPre, 2008; Ringwalt et al., 2003). In fact, a high level of implementation increases not only the opportunity to obtain statistically significant differences but also the chance to produce strong benefits for the program participants (Durlak & DuPre). However, when a program is implemented with a high level of fidelity but fails to produce anticipated outcomes, the researchers may need to redesign it (Nelson et al., 2012).

When reviewing more than 500 quantitative studies, Durlak and DuPre (2008) found strong evidence that a high level of implementation affects the outcome of any study. They also reviewed 81 reports and found at least "23 contextual factors that influence implementation" such as teacher self-efficacy, compatibility of the program or intervention with the organization's mission and values, flexibility of the program, setting climate, professional development workshops, and availability of resources (p.327). Durlak and DuPre also reviewed five meta-analyses conducted in health and education fields and published between 1986 and 2005 (Derzon, Sale, Springer, & Brounstein, 2005; DuBois, Holloway, Valentine, & Cooper, 2002; Smith, Schneider, Smith, & Ananiadou, 2004; Tobler, 1986; Wilson, Lipsey, & Derzon, 2003) and concluded that high levels of FOI influenced program outcomes positively.

When a program is implemented correctly as described by its developers, researchers are better able to interpret relationships between the program and its outcomes (Durlak, 1998). The data obtained from FOI are important for describing positive or negative outcomes. In some occasions, achieving negative outcomes may indicate that some core components of a program were not implemented in the right manner or were not implemented at all (Durlak). In addition, lack of fidelity may not affect the integrity of the program. For instance, in one national study measuring FOI in three educational and four criminal justice programs, "researchers reported that among the curricula that had been adapted, those with components that had been deleted or modified were less effective than those to which new components had been added" (Ringwalt et al., 2003, p.376).

Theoretical Framework

The research framework in this study was based on theories related to domain-specific expertise (Amabile, 1983; 1996; 2013; Sternberg, 1999). Amabile identified four components of her Componential Theory of Creativity, one of which is Domain-Relevant Skills (i.e., domain-specific expertise). This domain is considered a basic component of performance in any domain. Domain-specific expertise, in Amabile's theory, includes factors that influence the domain, such as an individual’s intelligence, the knowledge this individual gains about the particular domain, the individual expertise developed over the years, and the technical skills that are essential for the domain, such as laboratory techniques.

People who want to be creative problem solvers in a particular discipline need to have a deep understanding of relevant knowledge along with the ability to employ this knowledge in practice (Amabile, 1983; Hodges, 2005; Sternberg, 1999). In school, teachers are considered experts in their academic areas, thus, they have a responsibility to lead students toward developing their own knowledge, thinking abilities, and interpersonal skills (Hodges).

To develop domain-specific expertise, five elements should be considered: (a) “metacognitive skills”, (b) “learning skills”, (c) “thinking skills”, (d) “knowledge”, and (e) “motivation” (Sternberg, 1999). Metacognitive skills include an ability to understand a phenomenon, which includes the ability to define a problem and to find relevant resources to study the phenomenon in depth. Learning skills include using a systematic approach to acquiring the pertinent knowledge, such as distinguishing between relevant and irrelevant information that could be used to study the phenomenon and incorporating new knowledge into previous knowledge. Three main thinking skills contribute to developing expertise: (a) critical thinking skills (e.g., the ability to analyze, critique, and evaluate), (b) creative thinking skills (e.g., the ability to create a new element, invent a new element, and generate a new hypothesis), and (c) practical thinking skills (e.g., applying, employing, and practicing). Knowledge can be gained as declarative knowledge (e.g., facts and principles) or procedural knowledge (e.g., knowing how a phenomenon happens). Different kinds of motivation work as engines to keep individuals on track to becoming experts (Sternberg).

Significance of the Study

Preparing individuals to be creative problem solvers in science is a priority for any community. The global community constantly faces threat such as global warming, environmental pollution, and water and food shortages that may catastrophically affect future life on Earth. Developing students' creativity in science may contribute to reducing these dangers in the future (Kuo, 2016; NACCCE, 1999). In schools, science curricula should be designed in such a way that creativity among students is cultivated. One teaching model that can be implemented in any setting is the REAPS model (Maker & Zimmerman, 2008). One primary objective of this model is to increase creativity among students (Alhusaini et al., 2015; Maker et al., 2015; Wu et al., 2015). By conducting this study, the author had an opportunity to acquire evidence that supports or contradicts the claim that use of the REAPS model helped increase student creativity in the classroom.

One approach to examining the effectiveness of a certain teaching model is to determine its FOI (Azano et al., 2011). Until now, no study has been conducted to examine the relationship between FOI of the REAPS model and students' creative problem solving in science. Measuring FOI is necessary for increasing the confidence that the outcomes of the REAPS model come from the impact of implementing it in classrooms and not form other external variables. In fact, understanding whether a teaching model was implemented as intended by its developers allows researchers to interpret more accurately the association between the teaching model and observed results (Durlak, 1998; Gresham & Gansle, 1993; Moncher & Prinz, 1991). In many studies, the desired outcomes were mostly associated with a high level of implementation (Little, Riggs, Shin, Tate, & Pentz, 2015). However, a high level of implementation might not lead to achieving the desired results, in which case the teaching model would need to be redesigned.

By assessing FOI, this author is responding to the call of many governments and private agencies for more accountability in K-12 curriculum interventions research. In fact, by conducting this study, I obtained evidence-based research data about the relationship between the REAPS model and students' creative problem solving in science so that policymakers, schools principals, and teachers can decide whether to adopt this teaching model. In addition, the evidence-based research that comes out of this study may encourage teacher education programs to introduce this model to in-service and pre-service teachers for use in their classrooms.

The results of this study may allow researchers to identify which factors of the Test of Creative Problem Solving in Science (i.e., Finding Problems, Generating Detailed Solutions, and Classifying Elements) are most affected by the level of FOI of teachers. In addition, these results may include evidence of the association between creativity and developmental ages. The results that researchers reported in many studies on student creativity in elementary schools were mixed, especially among students in grades 4 and 6. In some of these studies, student creativity increased, while in the other studies their creativity declined.

Purpose and Research Questions

The purpose of this study was to determine the effect of the level of fidelity of implementation of the Real Engagement in Active Problem Solving (REAPS) model on creative problem solving in science on students from third to sixth grades. Specifically, the purpose was to examine the relationships between changes in students' growth in creative problem solving in science and the level of fidelity of implementation of the REAPS model. The following questions guided the study:

1. What were the differences in students' growth in creative problem solving in science at different grade levels?

2. What were the differences in growth of creativity of students who were in classrooms of teachers who were (a) at a high level of fidelity for two years (HF), (b) at a low level of fidelity for two years(LF), and (c) mixed levels of fidelity (i.e., high level of fidelity in Year 1 but low level of fidelity in Year 2 [HLF], and low level of fidelity in Year 1 but high level of fidelity in Year 2 [LHF])?
3. Which factors of the Test of Creative Problem Solving in Science (TCPS-S) were most affected by the level of fidelity of implementation of the teachers?

Definitions of Terms

The following terms can be defined theoretically and operationally in relation to the topic of this study:

Creativity in Science. The ability to solve a simple or complex scientific problem in the most efficient, effective, or economical ways, and these creative outcomes must be original to the relevant student age or audience.

Problem Solving. The process of answering questions, resolving difficulties, creating solutions, and investigating perplexing situations

Fluency. The ability to generate or produce many problems or solutions to a given question or situation (Bachelor & Michael, 1991; Guilford, 1959; Mumford & Gustafson, 1988).

Flexibility. The ability to sort many ideas into categories based on the similarities of the ideas (Bachelor & Michael, 1991; Guilford, 1959; Mumford & Gustafson, 1988).

Elaboration. The ability to add a meaningful description to an idea or to complete a creative idea (Bachelor & Michael, 1991; Guilford, 1959).

Originality. The ability to create or produce a new and novel idea or product (Bachelor & Michael, 1991; Guilford, 1959; Mumford & Gustafson, 1988).

The Real Engagement in Active Problem Solving (REAPS) Model. A teaching model developed by Maker and Zimmerman in 2004 to improve student ability to solve real-world problems in effective ways. It is a combination of three teaching models: the Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses (DISCOVER), Thinking Actively in a Social Context (TASC), and Problem-Based Learning (PBL).

The Test of Creative Problem Solving in Science (TCPS-S). A scientific test developed by Maker, Zimmerman, Pease, and Janes (2012), which includes two separate tasks: Task 1, Problems and Solutions, which consists of three sub-tasks (a) Identifying Problems, (b) Generating Solutions, and (c) Designing a Solution; and Task 2, Classifying Elements.

Fidelity of Implementation (FOI). The degree to which implementers such as teachers apply a program or teaching model in a manner that is described by the program developers (Dumas et al., 2001; Dusenbury et al., 2003; Mowbray et al., 2003).

The Research Team. A group of three members: a Principle Investigator (PI) who was a professor in education for gifted students for more than 40 years, a scientist who has worked with the PI for more than 14 years, and a former classroom teacher who had experience working with gifted students for more than 30 years and has worked with the PI for more than 25 years. The research team implemented the REAPS model in one elementary school in Australia. They trained the teachers in this school, reviewed and approved the teaching units that teachers developed, visited classrooms to assess the fidelity of implementation, and interviewed teachers, students, and parents to obtain their opinions of this model.

The Author. The author of the current study was one of the PI’s graduate students. He used the data that the research team had gathered during implementing the REAPS model in one elementary school in Australia. He participated in neither the data collection nor the assessment of teachers' levels of fidelity of implementation.

Chapter Summary

Problems and challenges in our world have increased significantly in the last few decades such as pollution, climate change, and natural resource depletion. Thus, creative minds in varied disciplines including science are required to solve these problems. In many studies, researchers emphasized the importance of providing specialized learning opportunities to students at early ages to increase their creativity. The REAPS model is one teaching model that can be implemented in any educational setting to increase creative problem solving among students.

In various studies, researchers pointed out the positive outcomes of the REAPS model for students. However, no study has yet been conducted to examine the association between the level of FOI and students' creative problem solving in science. In fact, linking the outcomes of the REAPS model with FOI is necessary to increase the confidence of the results and to respond to the trend of evidence-based research.

In this study, the author sought to measure the effect of the level of FOI of the REAPS model on creative problem solving in science on students from third to sixth grades. Specifically, the purpose is to examine the relationship between changes in student growth in creative problem solving in science and the level of FOI of the REAPS model.

CHAPTER II: REVIEW OF THE LITERATURE

The purpose of this chapter was to achieve two goals. The first was to evaluate the quality of studies conducted on Problem-Based Learning (PBL), Thinking Actively in a Social Context (TASC), the Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses (DISCOVER) curriculum model, and the Real Engagement in Active Problem Solving (REAPS) model. The second goal was to investigate the effect of the use of the PBL, TASC, DISCOVER, and REAPS models on students' creativity in science and to synthesize the major findings. General creativity, creativity in mathematics, and academic achievement were included in this chapter because a few researchers investigated the growth of creativity in science using these models. The questions that guided this review were as follow:

1. To what extent did studies on PBL, TASC, DISCOVER, and REAPS meet the quality indicators that were modified from the Council for Exceptional Children (2008)?
2. To what extent did students' general creativity, creativity in science and mathematics, and academic achievement increase after they were exposed to the PBL, TASC, DISCOVER, and REAPS models?

The chapter is divided into four sections. In the first section, the influence of the use of PBL on students' creativity in science and mathematics, general creativity, and academic achievement is discussed. The second section includes an examination of the impact of the use of TASC on students' creativity in science and mathematics and students' attitude and motivation for learning. In the third section, the influence of the use of the DISCOVER curriculum on students' creativity in science and mathematics and general creativity is discussed. Investigation of the impact of the use of the REAPS model and its effectiveness on students' creativity in science, general creativity, and students' perception of this model is examined in the fourth section.

Quality Indicators

The quality indicators developed by the Council for Exceptional Children (CEC) in 2008 were used and modified because these indicators were inclusive and covered quantitative and qualitative research designs. However, the mixed-methods design was not included by the CEC, so the elements of quantitative and qualitative indicators were combined to develop quality indicators for mixed-methods research.

The seven quality indicators for quantitative research were research conceptualization, settings and participants, intervention, fidelity of implementation (FOI), attrition, measurements, and data analysis. The six quality indicators for qualitative research were research conceptualization, settings and participants, intervention, data sources and analysis, trustworthiness and credibility, and outcomes. The eight mixed-methods research indicators included research conceptualization, settings and participants, intervention, fidelity of implementation (FOI), measurements, data analysis, trustworthiness and credibility, and outcomes. Each indicator for quantitative, qualitative, and mixed-methods research was comprised of two to seven sub-indicators.

A 5-point scale was used to evaluate each sub-indicator. One score was given if a sub-indicator was inadequate or unknown. Two scores were given if a sub-indicator was included but had no details. Three scores were given if a sub-indicator was partially stated. Four scores were given if a sub-indicator was almost adequately stated. Five scores were given if a sub-indicator was adequately stated.

Method for Locating Studies Reviewed in this Chapter

To find relevant studies to serve in a literature review, a researcher needs to develop a systematic and logical approach. In this review, three phases were included. The first phase was to develop criteria to include only relevant studies that reflect the purpose of the current study. The next phase was to create keywords to narrow the searching process. The final phase was to follow four steps to find studies: a keyword search, a review of the tables of contents in relevant scientific journals, a review of the reference lists in relevant articles, and a review of the relevant literature review and meta-synthesis studies. A summary of the method for locating studies reviewed in this chapter is presented in Figure 4.

Criteria for Selecting Studies Reviewed in this Chapter

After an initial reading of studies that were conducted on PBL, TASC, DISCOVER, and REAPS, five criteria for selecting eligible studies were developed. The importance of developing the criteria at this phase was to avoid deviation from the purpose of the current study and also to exclude studies that would not serve this review. First, studies published from 1985 to 2018 were chosen. The reason for narrowing the search to include only these dates was that the PBL, TASC, DISCOVER, and REAPS models were introduced or first developed in education during these years. PBL has been implemented in K-12 settings in the United States science the 1990s (Tan, 2003), so 1990 was the earliest date of PBL studies used in this research. TASC was first developed by Belle Wallace and Harvey Adams in 1985 at the University of Natal, South Africa (Wallace, 2008). DISCOVER was developed by C. June Maker in 1987 (Maker, 2001). The REAPS model was developed by C. June Maker and her colleagues in 2004 (Maker & Zimmerman, 2008). Second, only empirical studies in journal articles and dissertations were

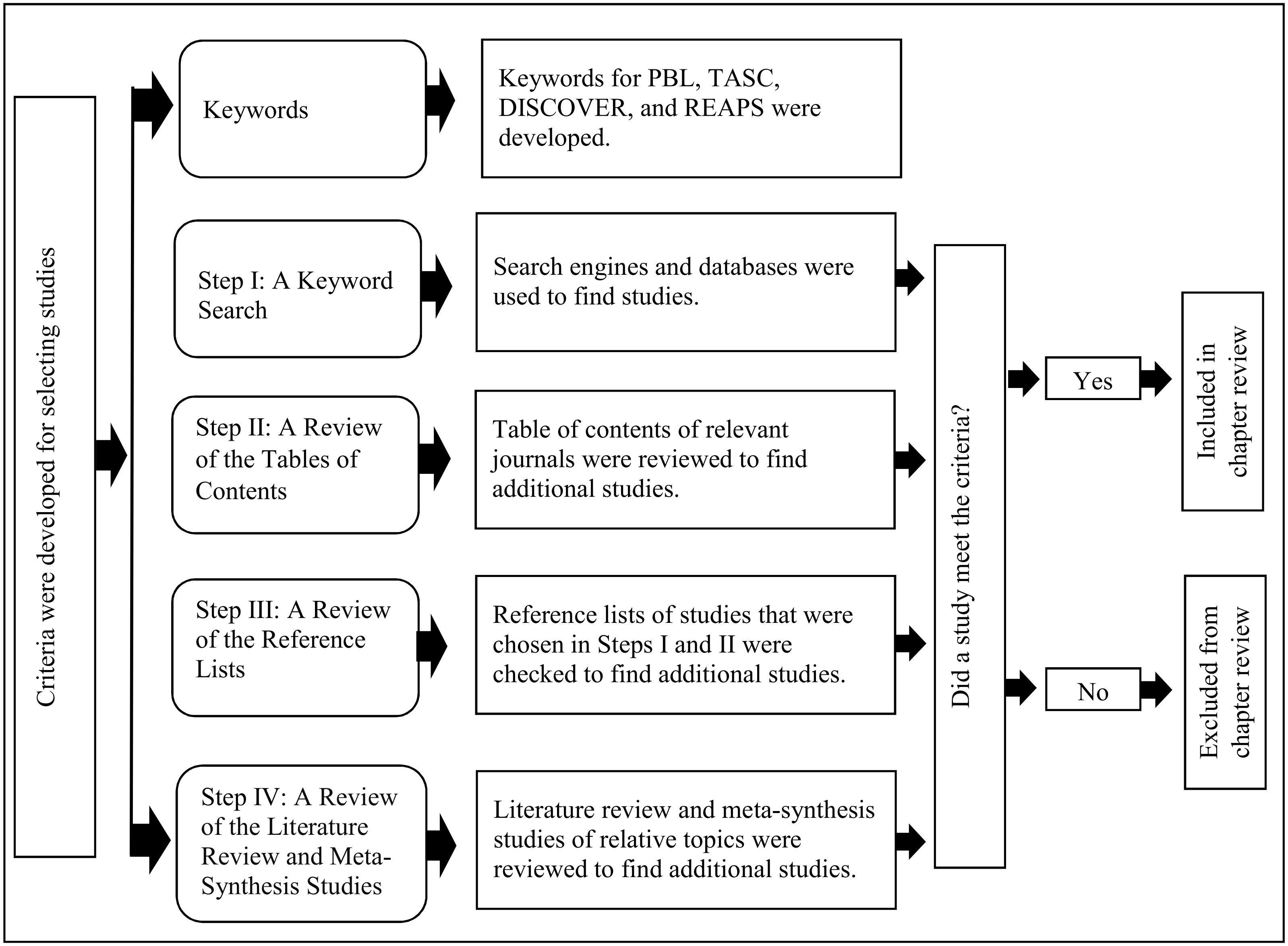

Figure 4: Process of Selecting Studies for Literature Review

included in this review. Thus, other resources such as books, theoretical and conceptual articles, studies based on personal experience without scientific evidence, and blogs were excluded. Third, only studies published in the English language were included. Fourth, only studies that were conducted in K-12 settings were used in this review. Fifth, studies that were conducted to examine the impact of the use of PBL, TASC, DISCOVER, and REAPS on students' creativity in science or mathematics, general creativity, and academic achievement were included.

Keywords

To select evidence-based research, two keyword categories were developed. The first category included primary terms such as *Real Engagement in Active Problem Solving* and its acronym, *REAPS*; *problem-based learning* and its acronym, *PBL*; *Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses* and its acronym, *DISCOVER*; and *Thinking actively in a Social Context* and its acronym, *TASC*. The second category included secondary terms: *science*, *scientific*, *mathematics, mathematical, creative*, *creativity*, and *academic achievement*.

Steps for Finding Studies

A systematic process was developed to locate studies for this review. The author followed four steps to find relevant studies.

Step I: A keyword search. Search engines were used to find relevant studies such as Primo, EBSCOhost, and Google Scholar with different educational databases such as ERIC, PsycINFO, Education Full Text, E-Journals, Teacher Reference Center, Professional Development Collection, and Academic Search Ultimate. I used search techniques for locating qualified studies such as using a hyphen. For example, searching for the term "problem-based learning" using a hyphen yields more results than searching for this term without a hyphen ("problem based learning").

During this step, article titles, abstracts, and keywords were read. The keywords developed in the first and second categories had to be in at least one of the article keywords, the abstract, or the title to be eligible for further review. However, if a study included keywords from only the second category without any keyword from the first category, the study was excluded. For example, "problem-based learning" was a primary term during my search on studies that were conducted on the PBL model, and "creativity" and/or "science" were secondary terms. Thus, any studies that contained only secondary terms were excluded from the review.

Hundreds of studies' titles and abstracts were read, then only relevant studies were selected. For example, 507 studies on PBL were found. However, after reading their abstracts carefully to ensure that they met this study's criteria, 13 studies were selected. The total number of studies that were chosen across PBL, TASC, DISCOVER, and REAPS was 29.

Step II: A review of the tables of contents. A list of relevant journals in the field of giftedness and creativity, three specialized journals about PBL, and journals in science education was made. The titles of these journals and the number of studies found are shown in Tables 2, 3, and 4. The purpose of reviewing the tables of contents of these journals was to find studies that might be missed in the keyword search (Step I). Each journal was reviewed, and the author read a table of contents for each issue starting from 1985 or the first issue if a journal was established after that year. This strategy was helpful, and more articles were found. Most of the studies found were conducted to examine the effect of using PBL on students' creativity, and no study was found of either TASC or REAPS. In total, nine studies were included during this step.

Table 2: *List of Specialized Journals about Giftedness and Creativity*

Title of Journal	Year		Articles Found			
	From	To	P	T	D	R
Gifted Child Quarterly	1985	2018	0	0	0	0
Journal of Creative Behavior	1985	2018	0	0	0	0
Journal for the Education of the Gifted	1985	2018	0	0	0	0
Gifted Child Today	1985	2018	0	0	0	0
Gifted and Talented International	1985	2017	0	0	0	0
Gifted Education International	1985	2018	0	0	0	0
Roeper Review: A Journal of Gifted Education	1985	2018	0	0	0	0
Creativity Research Journal	1988	2018	0	0	1	0
Journal of Secondary Gifted Education*	1995	2006	0	0	0	0
Journal of Advanced Academics	2006	2018	1	0	0	0
Thinking Skills and Creativity	2006	2018	0	0	0	0
Talent Development and Excellence	2009	2017	0	0	0	0
Creativity Studies	2008	2017	0	0	0	0
The New Zealand Journal of Gifted Education	2009	2015	0	0	0	0
Turkish Journal of Giftedness and Education	2011	2018	0	0	0	0

Note: P = Problem-Based Learning (PBL), T = Thinking Actively in Social Context (TASC), D = Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses (DISCOVER), R = Real Engagement in Active Problem Solving (REAPS).
* The name of this journal was changed to the Journal of Advanced Academics in 2006.

Table 3: *List of Specialized Journals about Problem-Based Learning*

Title of Journal	Year		Articles Found
	From	To	in PBL
Interdisciplinary Journal of Problem-based Learning	2009	2018	1
Journal of Problem Based Learning in Higher Education	2013	2018	0
Journal of Problem-Based Learning	2014	2018	0

Note: The Interdisciplinary Journal of Problem-based Learning was established in 2009 as the first specialized journal focused on problem-based learning.

Table 4: *List of Specialized Journals about Science Education*

Title of Journal	Year		Articles Found			
	From	To	P	T	D	R
Research in Science Education	1985	2018	1	0	0	0
Science Education	1985	2018	0	0	0	0
Studies in Science Education	1985	2018	0	0	0	0
International Journal of Science Education	1985	2018	0	0	0	0
Educational Research	1985	2018	0	0	0	0
Journal of Science Education and Technology	1992	2018	1	0	0	0
Journal of Science Education	2000	2008	0	0	0	0
Journal of Baltic Science Education	2002	2018	4	0	0	0
Journal of Technology and Science Education	2011	2018	0	0	0	0

Note: P = Problem-Based Learning (PBL), T = Thinking Actively in Social Context (TASC), D = Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses (DISCOVER), R = Real Engagement in Active Problem Solving (REAPS).

Step III: A review of the reference lists. The reference lists of the 38 studies selected during Steps I and II were reviewed to increase confidence that no valuable studies were missed. More than 1400 references among the 38 studies were checked, and five new PBL and DISCOVER studies were found. However, no new studies of TASC or REAPS were included at this step.

Step IV: A review of literature review and meta-synthesis studies. More studies could possibly be found if the primary purpose were to review literature in specific domains. In total, seven literature review and meta-synthesis studies of PBL were found, and 527 articles in the reference lists of the seven studies were checked. Only one new study was found. However, no literature review or meta-synthesis studies about TASC, DISCOVER, or REAPS were published. Table 5 includes a summary of the numbers of studies found during Steps I through IV.

Table 5: A *Summary of the Numbers of Studies Found during Step I thought IV.*

Step	Number of Studies Found				Total	Percentage
	P	T	D	R		
Step I	13	4	4	8	29	65.9%
Step II	8	0	1	0	9	20.4%
Step III	4	0	1	0	5	11.4%
Step IV	1	0	0	0	1	2.3%
Overall	**26**	**4**	**6**	**8**	**44**	**100%**

Note: P = Problem-Based Learning (PBL), T = Thinking Actively in Social Context (TASC), D = Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses (DISCOVER), R = Real Engagement in Active Problem Solving (REAPS).

Problem-Based Learning (PBL)

This section contains an overall summary of studies that were conducted on the PBL model. The results of evaluation using the quality indicators and investigation of the effect of the use of PBL on students' general creativity, creativity in science and mathematics, and academic achievement are discussed.

A Summary of PBL Studies

The PBL model was used as an intervention in 26 studies to measure its effect with students in classrooms. The authors of these studies implemented different research designs: 20 quantitative studies, three qualitative studies, and three mixed-methods design studies.

The researchers who used a quantitative research design conducted their studies in six countries. Six studies were implemented in Indonesia, five in the USA, three each in Malaysia and Turkey, one study each in South Africa and South Korea, and Dods (1997) did not state the study location. Two qualitative studies were implemented in Indonesia, and one study was conducted in the USA. The three mixed-methods studies were implemented in three countries: Hong Kong, Malaysia, and Turkey.

The PBL model was inclusive and covered all K-12 settings. Eleven studies were implemented in high schools, nine in middle schools, three each in elementary schools and kindergartens. Most of the studies (80.8%) were conducted in urban areas, 3.8% in suburban areas, and the researchers in 15.4% of studies did not specify the area in which the research was conducted. The classroom size in 50% of the studies was 30 students or more, while only 7.7% of the studies were implemented in classrooms that had fewer than 20 students. Five studies were conducted with gifted students, while the classroom type in 73.1% of studies was not reported.

Teachers in 42.3% of the studies participated in professional development workshops to implement PBL in the classroom, while no information was stated about the teachers in the rest of the studies receiving any professional development workshops. The subject teachers taught in all studies was science, and researchers in only two of the studies combined science with mathematics or social studies. Teachers' experience, which ranged from one to more than ten years, was stated in only seven studies. In addition, teachers' gender was not reported in 21 studies.

The authors of 76.9% of studies used more than one instrument to gather their data. The common methods of collecting data were achievement tests, observations, and creativity tests developed by the researchers, such as the Figural Scientific Creativity Test. The researchers used mostly t-tests and Multivariate Analysis of Variance (MANOVA) to analyze their data. Appendix A has a descriptive classification of these studies of PBL.

The Results of the Evaluation of Quality of PBL Studies

In this section, the results of the evaluation of quality of PBL studies are discussed. They are divided based on research designs: quantitative, qualitative, and mixed-method.

Quantitative studies. The researchers in 20 studies used this design to investigate the impact of implementing PBL on students in classrooms. Seven indicators (CEC, 2008) were used to examine the quality of the studies. These indicators included research conceptualization, settings and participants, intervention, fidelity of implementation (FOI), attrition, measurements, and data analysis. The study by Horak and Galluzzo (2017) scored highest, including almost all quality indicators and receiving a total score of 95.9%, while the study by Ratnasari, Supriyanti, and Rosbiono (2017) scored lowest and earned only 36%. The other 18 studies had percentages ranging from 51% to 84.1%. The average percentage across all 20 studies was 73%. The scores received for each quality indicator, its sub-indicators, and the total score for each study are presented in Appendix B.

For research conceptualization, which was comprised of four sub-indicators, full scores (five) were given for all studies on two sub-indicators, rationale and purpose of the study. Authors discussed the importance of their research and the purpose for conducting their studies. For example, Anazifa and Djukri (2017) emphasized that the importance of teaching creativity and critical thinking to students in classrooms led them to conduct their research, and that the purpose was to examine the impact of the use of project-based learning and problem-based learning on students' creativity and critical thinking. However, in the other two sub-indicators of the research conceptualization, theoretical framework and research questions, the scores received by studies were varied, ranging from one to five. For instance, in 30% of the articles, the theoretical framework was not discussed (Araz & Sungur, 2007; Jo & Ku, 2011; Nurdin & Setiawan, 2015; Ratnasari et al., 2017; Sungur, Tekkaya, & Geban, 2006; Wartono, Diantoro, & Bartlolona, 2018), and the research questions in 40% of articles were not stated clearly and were only implied in the purpose statements (Aidoo, Boateng, Kissi, & Ofori, 2016; Anazifa & Djukri, 2017; Gordon, Rogers, Comfort, Gavula, & McGee, 2001; Jo & Ku, 2011; Mundilarto, 2017; Nurdin & Setiawan, 2015; Ratnasari et al., 2017; Wartono et al., 2018).

The settings and participants are considered the heart of each study, and researchers should provide details needed to allow future researchers to duplicate their research. This indicator was comprised of five sub-indicators: (a) a description of the setting, (b) a description of how the setting was selected, (c) appropriate procedures for selecting the participants, (d) relevant demographic information about the participants, and (e) a description of the researcher's role in relation to the settings and participants. However, full scores were given to only 11.5% of the studies across all five sub-indicators (Gallagher & Gallagher, 2013, Horak & Galluzzo, 2017, Wong & Day, 2009), while the researchers in one study earned low scores, ranging from one to two on these sub-indicators (Nurdin & Setiawan, 2015).

The average scores earned across the 20 studies on five out of six sub-indicators in the intervention category ranged from 4 to 4.8 (out of 5), which was one of the highest scores among all indicators. The researchers in general met an acceptable level of quality on the following sub-indicators: (a) interventionists' roles, (b) interventionists' conditions, (c) a description of PBL, (d) a description of how PBL was implemented, and (e) relevant aspects of practice in the control group. However, the demographic information for the interventionists was reported in only four studies (Gallagher & Gallagher, 2013; Horak & Galluzzo, 2017; Siew & Mapeala, 2016; Wong & Day, 2009).

One crucial quality indicator researchers should consider is assessing the fidelity of implementation (FOI) of the teaching model. Unfortunately, this indicator was ignored by most researchers in the PBL studies. The average scores received in the 20 studies ranged from 1.2 to 1.8 (out of five) on the three sub-indicators: (a) FOI was assessed, (b) key features of practice were assessed, and (c) inter-observer reliability of FOI was collected regularly. Researchers in only three of the studies stated that they assessed FOI (Horak & Galluzzo, 2017; Siew & Chin, 2018; Siew, Chin, & Sombuling, 2017). However, Siew and Chin and Siew et al. did not describe how they assessed FOI in their research.

For the attrition indicator, almost all of the scores on its two sub-indicators were 5. The possible reason for this indicator receiving the highest scores among the seven indicators was the fact that the duration of these studies ranged from 2 weeks to 6 months, which reduced the possibility that the participants left their schools during these times.

The researchers of 15 studies employed more than one instrument to collect their data, which may help to increase the credibility of the interpretations of their findings. In addition, in 75% of the studies, the researchers described in detail how they collected their data. However, the validity of instruments was reported in only 20% of the studies, which included concurrent, content, construct, and/or predictive validity; in 15% of the studies, the outcome measures had evidence of internal consistency; and in 40% of the studies, the effect size value was reported.

Qualitative studies. The researchers in three studies used a qualitative design to examine the effect of the use of PBL in classrooms. To investigate the quality of these studies, six indicators were implemented (CEC, 2008): research conceptualization, settings and participants, intervention, data sources and analysis, trustworthiness and credibility, and outcomes. In general, the average percentages of quality indicators achieved among these studies ranged from 49.1% (Nurdyani, Slamet, & Sujadi, 2018) to 80% (Zhang, Parker, Eberhardt, & Passalacqua, 2011). The average percentage across all studies was 68.7%. The scores received for each quality indicator, its sub-indicators, and the total score for each study are presented in Appendix B.

For research conceptualization, which included (a) rationale for the study, (b) purpose of the study, (c) theoretical framework, and (d) research questions, the average score received by studies across the four sub-indicators was 4 (out of 5). Full scores were given on the rationale and purpose of study sub-indicators across all studies. The researchers justified their research and the purpose of implementing their studies. For instance, Haridza and Irving (2017) highlighted the importance of changing the way students acquired knowledge by giving chances for students to be active learners through PBL instead of using direct teaching in classrooms. However, Haridza and Irving were the only researchers who described the theoretical framework for their research.

Similar to the quantitative research, the settings and participants indicator was comprised of five sub-indicators. On average, a score of 3.6 was given to the three qualitative studies on all five sub-indicators. Although two studies received almost full scores across all the sub-indicators, Nurdyani et al. (2018) had the lowest scores, ranging from 1 to 2, which affected the overall average. In this study, for example, the researchers did not use appropriate procedures for selecting students to participate in their research. They presented problems that were covered in the coordinate systems teaching unit, and students who received 80 points out of 100 were selected as gifted students in mathematics.

The intervention indicator had one of the highest average scores (4.5) across the six indicators. The researchers in only one study (Nurdyani et al., 2018) failed to report the interventionists' roles while implementing PBL in classrooms, although they provided an overall explanation of the PBL model in the introduction to their research. On the other hand, demographic information about the teachers who implemented PBL, their roles in classrooms, and a description of how PBL was implemented in classrooms were stated in detail in the other two studies.

The lowest scores given were in the categories of data sources and analysis and trustworthiness and credibility. The average scores given ranged from 1 to 2.46. The researchers, in general, did not provide sufficient information about these indicators. For example, Nurdyani et al. (2018) provided no information about who analyzed the data and how the researchers established trustworthiness and credibility when interpreting their results. Haridza and Irving (2017) conducted interviews with students but did not illustrate how they dealt with the interview data in their findings. Only one sub-indicator of data sources and analysis was given full scores across all three studies. All researchers used more than one instrument to collect their data. However, none of the authors provided any information about trustworthiness and credibility.

High scores were given to the two sub-indicators of the outcomes category, and the average score was 4.67. In all studies, researchers provided a clear and detailed description of the outcomes of their studies, and they highlighted the strengths of their findings.

Mixed-methods studies. The authors of three articles employed a mixed-methods design to explore the effect of the use of PBL with students in classrooms. To examine the quality of these studies, eight indicators were used (CEC, 2008): research conceptualization, settings and participants, intervention, fidelity of implementation (FOI), measurements, data analysis, trustworthiness and credibility, and outcomes. Each indicator was comprised of sub-indicators ranging in number from two to five. In general, the average percent of scores for the three studies using mixed-methods designs ranged from 45.7% (Fatimah, 2015) to 84.3% (Siew, Chong, & Lee, 2015). The average score across the three studies was 66.7%. The scores received for each quality indicator, its sub-indicators, and the total score for each study are presented in Appendix B.

The research conceptualization indicator included (a) rationale for the study, (b) purpose of the study, (c) theoretical framework, and (d) research questions. The average score received by studies across the four sub-indicators was 4.42 out of 5. Full scores (5) were given to all studies on the sub-indicators rationale and purpose of the study, and only Fatimath (2015) received a score of 1 on the research question sub-indicator. She did not provide any questions.

For the settings and participants category, only one study had scores of 5 across all five sub-indicators. These sub-indicators were (a) a description of the setting, (b) a justification of the setting selection, (c) appropriate procedures for selecting the participants, (d) relevant demographic information about the participants, and (e) a description of the researcher's role in relation to the settings and participants (Siew et al., 2015). The average score given for this indicator was 3.47 for the three studies.

The second highest score given across the other seven indicators was for the intervention category. However, a score of only 1 was given to the relevant demographic information about interventionists sub-indicator across all three studies. The researchers failed to provide any information about the teachers who implemented PBL in their classrooms.

In contrast to intervention, FOI had the lowest scores across all indicators. The researchers did not assess FOI in classrooms. The failure to assess FOI might raise questions about the findings of these three studies because FOI is considered an essential element for any study whose researchers seek to attribute the impact of their intervention on students to the intervention itself rather than to other external variables.

One strength in the measurement category was that all researchers described how the data were collected, and used more than one instrument to gather information about implementing PBL in classrooms. However, Akınoğlu and Tandoğan (2007) were the only researchers who reported the validity and reliability of the three instruments, which included an achievement test, an open-ended questions test, and an attitude scale for science education.

For the data analysis, trustworthiness and credibility, and outcomes categories, the average scores received by studies ranged from 2.3 to 3.67. Fatimath (2015) discussed neither the method of data analysis nor how she implemented trustworthiness and credibility to increase the fidelity of interpreting the results of the study. In contrast, Siew et al. (2015) provided a detailed discussion of their data analysis. They stated that a science teacher and one researcher scored students' answers independently. Both of the raters attended a professional development workshop to learn the method for scoring students' responses. In the three studies, the effect size was not reported.

The Effect of the Use of PBL on Students' Creativity and Academic Achievement

Problem-based learning is a popular teaching method that has been implemented in a variety of contexts and disciplines such as medicine, engineering, and education (Yew & Goh, 2016). Hundreds of PBL studies have been conducted in the field of education. However, few researchers have examined the effect of the use of PBL on students' general creativity, creativity in science and mathematics, and academic achievement. For example, in the *Interdisciplinary Journal of Problem-based Learning*, which is focused on PBL studies, only one study among 128 (0.78%) that were published from 2009 to 2018 was focused on general creativity or creativity in science. Another example is a literature review conducted by Merritt, Lee, Rillero, and Kinach (2017). Among 504 studies that Merritt et al. found, researchers in only nine studies examined the effectiveness of PBL in mathematics and science education in K-8 settings.

Twenty-six studies were qualified for his review. To report the results of these studies and discuss the effect of the use of PBL with students in classrooms, these studies were categorized based on their purposes. In general, three purposes were identified among the 26 studies: (a) PBL and its impact on students' creativity and critical thinking skills when PBL was the only intervention, (b) PBL and its influence on students' academic achievement when PBL was the only intervention, and (c) PBL and its effect on students' creativity and academic achievement when it was combined with or compared to other teaching methods, such as project-based learning and cooperative learning.

Researchers in nine studies examined the influence of the use of PBL on students' creativity and critical thinking skills. However, they explored the effect of PBL in different ways. For example, high school was the setting for five studies implemented in Indonesia (Fatimah, 2015; Ratnasari et al., 2017), South Africa (Aidoo et al., 2016), the USA (Gallagher & Stepien, 1992), and South Korea (Jo & Ku, 2011), while middle school was the setting for two studies implemented in Indonesia (Haridza & Irving, 2017; Nurdyani et al., 2018), and elementary school was the setting for two studies in Malaysia (Siew et al., 2015; Siew & Mapeala, 2016). The researchers in eight studies found a positive impact on students' creativity and critical thinking after using PBL as an interventuon in classrooms. For instance, Aidoo et al. implemented PBL in five high schools and used a standardized test to gather their data. They concluded that PBL was an effective method to teach chemistry. In the treatment group, the students scores on critical thinking and problem-solving tests were higher than those of the control group. Another example was that Haridza and Irving found students' critical thinking skills had increased sharply after the teachers implemented PBL in science classes. Only Siew and Mapeala found that students' scientific creativity scores were higher in classrooms in which the teacher combined PBL with a Thinking Maps strategy to teach the students, compared with students whose teachers used only PBL or traditional teaching methods. However, students' scientific creativity scores in the PBL class were higher than those of students in the traditional teaching methods class.

In another ten studies, the researchers investigated the influence of the use of PBL on students' academic achievement. PBL was used as an intervention, and its impact on students' academic achievement and understanding of science content was examined in high school (Mundilarto, 2017; Sungur et al., 2006), middle school (Akınoğlu & Tandoğan, 2007; Araz & Sungur, 2007; Gallagher & Gallagher, 2013; Gordon et al., 2001; Horak & Galluzzo, 2017; Inel & Balim, 2010; Wong & Day, 2009), and kindergarten (Zhang et al., 2011). In the ten studies, PBL had a crucial role in improving students' academic achievement and their attitudes toward learning. For example, Akınoğlu and Tandoğan conducted their study on 50 students in grade 7 in one middle school. The researchers used an achievement test, an open-ended questions test, and an attitude scale for science education to examine the effect of the use of PBL. By the end of the intervention, students' academic achievement and attitudes toward science improved. Sungur et al. concluded that students who were instructed with PBL scored higher in both performance skills and academic achievement than did students who were instructed with traditional teaching methods. PBL also had a positive impact on academic achievement among low SES level students (Gallagher & Gallagher; Gordon et al., 2001). Gallagher and Gallagher concluded that PBL played an important role in encouraging low income-students to express their academic potential.

Problem-based learning also was combined with or compared to other teaching models to examine their influence on students' creativity and academic achievement (Anazifa & Djukri, 2017; Dods, 1997; Drake & Long, 2009; Wartono et al., 2018; Nurdin & Setiawan, 2015; Siew & Chin, 2018; Siew et al., 2017). PBL was integrated with web-based e-learning and cooperative learning or compared with project-based learning and traditional teaching methods. The results were mixed. For example, Anazifa and Djukri conducted a study among high school students in three classrooms to examine the difference between project-based learning and problem-based learning. They used three instruments to gather their data: observation, a creativity test, and a critical thinking test. The researchers found that both project-based learning and problem-based learning were effective teaching models for improving students' creativity and critical thinking skills. They found that the creativity scores of students in the project-based learning classroom were higher than those in the problem-based learning classroom. However, the results of students' critical thinking tests were not statistically significant among these classrooms.

Dods (1997) implemented his study in one high school, investigating the difference between the use of PBL and traditional teaching methods on students' understanding of science content in one high school. He concluded that the use of PBL promoted in-depth understanding of knowledge, while traditional lecture methods played an essential role in content coverage. In contrast, Drake and Long (2009) conducted their research on students in an elementary school. They found that the scores in science content knowledge of the students in the treatment group, who were taught with PBL, were higher than those of students in the control group. Furthermore, Siew et al., 2017 concluded that students in a class in which PBL was combined with cooperative learning scored higher than did students in a class that was taught using only PBL or in a class that was taught using traditional methods; these same students also had higher scores in scientific creativity, as measured by fluency, elaboration, originality, abstractness of title, and resistance to premature closure. However, students in the PBL class scored higher than did students in a traditional class in scientific creativity.

Although the researchers in all 26 PBL studies concluded that the use of this model helped increase students' creativity and academic achievement, the quality of six studies (Anazifa & Djukri, 2017; Fatimah, 2015; Gordon et al., 2001; Nurdyani et al., 2018; Ratnasari et al., 2017; Wartono et al., 2018) are questionable. The scores in my evaluation of each of them were less than 60%. To improve confidence in the results of their studies, the researchers should have explained the process they used to select participants and the ways PBL was implemented. In addition, the instruments used may have lacked validity and reliability. Data on these qualities of the instruments were not reported. One of the important research aspects to which the researchers did not pay attention was assessing FOI among the teachers in classrooms. None of these researchers observed the teachers to ensure they implemented PBL in an effective manner. A summary of each study of PBL is presented in Table 6.

Table 6: *A Summary of PBL Studies*

Author(s)	Participants	Settings	Research Design	Purpose
Aidoo et al., 2016	• 102 students	• Five high schools	• Quasi-experimental	• Examined the impact of the use of PBL on students' achievement in chemistry.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Standardized test	• T-test	• Science	• The researchers concluded that PBL was an effective method to teach chemistry. Students' critical thinking and problem-solving skills in the treatment group were increased more than those of students in the control group.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Akınoğlu & Tandoğan, 2007	• 50 students in grade 7	• One middle school	• Mixed-methods	• Investigated the influence of the use of PBL on students' academic achievement and concept learning in science.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Achievement test • Open-ended questions • Attitude scale for science education.	• T-test • Coding and developing themes	• Science	• The researchers concluded that PBL had a positive effect on students' academic achievement and attitudes toward science.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Anazifa & Djukri, 2017	• 102 students in grade 11	• Three classrooms in high school	• Quasi-experimental	• Examined the impact of the use of project-based learning and problem-based learning on students' creativity and critical thinking.

(Continued)

Table 6: *A Summary of PBL Studies*

	Data Source	Data Analysis	Subject(s)	Findings
	• Observation • Creativity test • Critical thinking test	• T-test • MANOVA	• Science • Mathematics	• Both project-based learning and problem-based learning were effective models to develop student's creativity and critical thinking. Students' creativity scores in the project-based learning group were higher than those of students in the problem-based learning group. However, no differences were found between students' critical thinking scores in both groups.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Araz & Sungur, 2007	• 217 students in grade 8	• One middle school	• Quasi-experimental	• Examined the influence of the use of PBL on students' academic achievement and performance skills in one middle school in a science unit.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Achievement test • The Test of Logical Thinking	• ANCOVA	• Science	• Scores in science units of students in the treatment group who were taught using PBL were higher (M = 11.44 and M = 2.67) than scores of students in the control group who were taught science units using traditional methods (M = 10.91 and M = 2.20).
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Dods, 1997	• 30 students	• One high school	• Quantitative-Action research	• Investigated which of three teaching models (PBL, a traditional lecture, and a combination of PBL and traditional lecture) was more effective to increase understanding and retention of the principal content in a science class.

(Continued)

Table 6: *A Summary of PBL Studies*

Author(s)	Data Source	Data Analysis	Subject(s)	Findings
	• Self-evaluations • Actual Depth of Understanding survey • Questionnaire	• ANOVA	• Science	• The researcher concluded that PBL contributed in-depth understanding of knowledge while a traditional lecture played an essential role in content coverage.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Drake & Long, 2009	• 33 students in grade 4	• One elementary school	• Quasi-Experimental	• Examined the effect of the use of PBL on students' knowledge in science and retention of information over time compared to students who were taught with the direct teaching approach.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Content knowledge • The Draw-a-Scientist Test • Interviews • Observations	• T-test	• Science	• The scores of students in the treatment group were higher than scores of students in the control group in content knowledge in science.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Fatimah, 2015	• 30 junior high school students	• One high school	• Quantitative-action research	• Investigated whether the use of PBL increased creativity and critical thinking skills in science among junior high school students through implementing PBL.

(Continued)

Table 6: *A Summary of PBL Studies*

	Data Source	Data Analysis	Subject(s)	Findings
	• Creative products. • Questionnaire • Interviews • Observations • Score tests of critical thinking	• Not clear	• Science	• Students' creativity and critical thinking skills scores increased by 67% and 74.4% respectively after integrating and implementing PBL with the Jelajah Alam Sekitar (JAS) approach.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Gallagher & Gallagher, 2013	• 271 sixth grade students • 13 teachers	• Two middle schools	• Quasi-experimental	• Explored whether the use of the PBL model could contribute to encouraging low income-students to demonstrate their academic potential. Also, the researchers aimed to investigate whether the use of PBL could identify gifted students from a low-income population.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Standardized achievement test • Teacher ratings • Independent ratings • PBL assignments	• ANOVA • Chi-square	• Science • Social studies	• The researchers concluded that PBL played an important role in encouraging students to express their academic potential and helping to identify gifted students from a low-income population.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Gallagher & Stepien, 1992	• 120 junior and senior high school students	• One high school for gifted students	• Quasi-experimental	• Investigated the effect of implementing PBL on students' problem-solving skills.

(Continued)

Table 6: *A Summary of PBL Studies*

	Data Source	Data Analysis	Subject(s)	Findings
	• Problem-solving test	• Chi-square	• science	• The researchers concluded that PBL was an effective model for improving student problem solving especially if one problem was presented to the students.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Gordon et al., 2001	• 196 students from grade 6 to 8	• One middle school	• Quasi-experimental	• Examined the effect of the use of PBL on students' achievement in science among low-income minority groups (African American and Hispanic) when they were exposed to PBL for a small percentage of total curriculum time.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Feedback • Concept maps • Self-assessment	• Not clear	• Science	• The researchers concluded that exposing students to PBL for a small percentage of time (2% of the total curriculum) improved their achievement.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Haridza & Irving, 2017	• 32 students in grade 8	• One middle school	• Qualitative-Action Research	• Investigated whether the use of PBL improved middle school students' critical thinking skills in science.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Self-assessment • Peer assessment • Rating scale • Checklist • Interview	• Not clear	• Science	• The researchers concluded that students' critical thinking skills had increased sharply after their teachers implemented PBL in science class.

(Continued)

Table 6: *A Summary of PBL Studies*

Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Horak & Galluzzo, 2017	• 449 gifted students in grade 7	• Two middle schools	• Quasi-experimental	• Investigated the impact of implementing PBL on students' achievement and their perceptions of classroom quality.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Standardized test in science • Self-rating survey	• T-test	• Science	• After implementation of PBL in the treatment class, students' academic achievement increased more than the achievement of students in the control group. The results were statistically significant. • Students in the treatment group expressed positive attitudes toward learning more than those of students in the control group.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Inel & Balim, 2010	• 41 students in grade 7	• One middle school	• Quasi-experimental	• Examined the influence of implementing PBL on middle school students' academic achievement after integrating PBL with a science and technology teaching unit.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Academic achievement test • Open-ended questions	• Mann Whitney U Test • The Wilcoxon Signed-Ranks Test	• Science	• The researchers concluded that PBL had a positive impact on students' academic achievement in the treatment group. The results were statistically significant.

(Continued)

Table 6: *A Summary of PBL Studies*

Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Jo & Ku, 2011	• 151 gifted students	• One high school	• Action research	• Examined the effectiveness of PBL with gifted students in one high school in science classes.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Creative self-report questionnaire • Self-regulation • Discussion	• T-tests	• Science	• The researchers found that participants' scores in self-regulation, creativity, the level of interest, and the frequency of discussions in posttests were higher than scores in pretests.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Mundilarto, 2017	• 64 students in Grade 10	• One high school	• Quasi-experimental	• Examined the effect of the use of PBL on students' adademic achievement and critical thinking among students in grade 10 in physics classes.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Physics achievement test. • Students' critical thinking test • Observation	• MANOVA	• Science	• After implementation of PBL in the treatment group class, students' academic achievement and critical thinking skills scores were higher than scores of students in the control group. The results were statistically significant.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Nurdin & Setiawan, 2015	• 48 in grade 9	• One high school	• Quasi-experimental	• Examined the effect of the use of web-E-learning with PBL on students' cognitive and critical thinking skills among students in a grade 9 science class.

(Continued)

Table 6: *A Summary of PBL Studies*

Author(s)	Data Source	Data Analysis	Subject(s)	Findings
	• Physics achievement test • Students' critical thinking test	• Normalized gain	• Science	• Students' cognitive and critical thinking skills scores in the treatment group whose teacher integrated the web-E-learning model with PBL in the science class were higher than those of students in the control group whose teacher used only PBL. The results were statistically significant.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Nurdyani et al., 2018	• 36 students in grade 8	• One middle school	• Qualitative descriptive research.	• Described the level of creative thinking in mathematics among gifted students in grade 8 after exposing them to PBL.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Observation • Tests • Interviews	• Fluency, flexibility, and originality criteria	• Mathematics	• The researchers pointed out that students' creative thinking in mathematics, measured by fluency, flexibility, and originality, was higher after their teachers implemented PBL.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Ratnasari et al., 2017	• 32 in grade 7	• One vocational high school	• Quasi-experimental	• Examined the effect of the use of PBL on students' creativity in a food additives course.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Essay • Student worksheet • Observation	• Calculating percentages	• Science	• The researchers concluded that creative thinking skills increased among the participants.

(Continued)

Table 6: *A Summary of PBL Studies*

Author(s)	Participants	Settings	Research Design	Purpose
Siew & Chin, 2018	• 144 six-year-old preschoolers	• Three kindergartens	• Quasi-experimental	• Evaluated the effectiveness of PBL and cooperative learning on preschoolers' scientific creativity.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• The Figural Scientific Creativity Test	• T-test	• Science	• The results in the treatment group whose teachers integrated PBL with Cooperative Learning were statistically significant than those of students in the control group, $t(142) = 6.73$, $p < .05$.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Siew et al., 2017	• 216 preschoolers-age 6	• Three kindergartens	• Quasi-experimental	• Examined the increase in scientific creativity among preschoolers who were exposed to either a PBL and cooperative learning model, or a PBL model only, or taught with traditional methods.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• The Figural Scientific Creativity Test	• MANOVA • MANCOVA	• Science	• Students' scientific creativity scores were higher in classrooms in which their teachers integrated PBL with a cooperative learning model compared to students whose teachers used only PBL or traditional teaching methods. Also, students' scientific creativity scores in the PBL class were higher than those of students in the traditional methods class.

(Continued)

Table 6: *A Summary of PBL Studies*

Author(s)	Participants	Settings	Research Design	Purpose
Siew et al., 2015	• 232 students in grade 5	• Two elementary schools	• Mixed-Methods	• Investigated whether the use of PBL increased scientific creativity among students in grade 5.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Open-ended questions • Scientific creativity tests	• T-test • Coding and categorizing data	• Science	• In the posttest, students scored higher in technical products, scientific knowledge, and scientific problems of the test components than they did in the pretest. Also, students reported that PBL was fun, easy, and interesting as well as a practical way of acquiring scientific knowledge.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Siew & Mapeala, 2016	• 270 fifth grade students	• Three elementary schools	• Quasi-experimental	• Examined the increase in scientific thinking skills among students in grade 5 who were exposed to PBL and Thinking Maps in a physical science class, only PBL in a physical science class, and traditional teaching methods in a physical science class.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• The Test of Science Critical Thinking	• MANOVA • MANCOVA	• Science	• Students' scientific creativity in the classes in which their teachers combined PBL with Thinking Maps scored higher than students in the classes in which their teachers implemented only PBL or used traditional teaching methods. Also, students' scientific creativity scores in the PBL classes were higher than those of students in the traditional methods classes.

(Continued)

Table 6: *A Summary of PBL Studies*

Author(s)	Participants	Settings	Research Design	Purpose
Sungur et al., 2006	• 61 students in grade 10	• One high school	• Quasi-experimental	• Investigated the influence of the use of PBL on students' academic achievement and performance in a human excretory system unit.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Achievement test • Problem based learning feedback form	• ANOVA • MANOVA	• Science	• Scores of the students who were taught using PBL were higher than those of students who were taught using traditional teaching methods in both performance skills and academic achievement.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Wartono et al., 2018	• 77 students in grade 10	• One high school	• Quasi-experimental	• Examined the influence of the use of PBL and traditional learning methods on students' creative thinking skills in science classes.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Creative Thinking Skills test	• ANCOVA	• Science	• Students' creative thinking skills scores in the PBL class were higher than those of students in the traditional teaching methods class.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Wong & Day, 2009	• 75 students	• One middle school	• Action research design	• Examined the influence of implementing PBL and lecture-based learning on students' achievement in science class.

(Continued)

Table 6: *A Summary of PBL Studies*

	Data Source	Data Analysis	Subject(s)	Findings
	• Multiple choice questions • Questionnaire	• T-test	• Science	• PBL and lecture-based learning had an almost similar effect on knowledge acquisition. However, the students in the PBL class showed improvement in application and comprehension of knowledge over an extended period of time.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Zhang et al., 2011	• One teacher and 24 kindergarten students	• One kindergarten	• Qualitative-Action Research	• Investigated the effectiveness of implementing PBL in kindergarten science class on students' learning and understanding and how the teacher applied PBL in her science class.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Research plan • PBL video lesson • Student assessment data • Teacher's notes • Teacher's final report • Summer evaluation • Surveys • Interviews	• Not clear	• Science	• The researchers concluded after analyzing the teacher's documents that students showed improvement in their content understanding. Also, the researchers pointed out that students' questioning skills improved.

Thinking Actively in a Social Context (TASC)

This section contained three parts. First, a summary of studies of TASC was reported. The quality of studies conducted to implement TASC in classrooms was discussed in the second part. Finally, the effects of the use of TASC on students' creativity in science, mathematics achievement, and attitude and motivation toward learning were reported.

A Summary of TASC Studies

Only four TASC studies were found that met the criteria in the current study. One was a quantitative design (Abu Awwad, Asha, & Jado, 2014), two were qualitative (Davies, 2008; Faulkner, 2008), and a mixed-methods design was used in the other study (West, 2008). Three studies were implemented in elementary schools in the UK, and one study was conducted in a Jordanian middle school. No information was provided about the school area: urban, suburban, or rural.

The classroom size in one study was fewer than 20 students, two studies were implemented in classrooms that included 30 students or more, while a researcher in one study did not report classroom size. One study was conducted with gifted students while the researchers in the other three did not state the classroom type. Among all studies, no information was given about teaching experience of the teachers or whether they attended professional development workshops to implement TASC in their classrooms. The researchers in two studies stated that science and mathematics were the subjects taught in classrooms. The number of instruments for collecting data ranged from two to four, and the common methods for gathering data were observations and questionnaires. Appendix A has a descriptive classification of these studies of TASC.

The Results of the Evaluation of Quality of TASC Studies

To report the results of the quality of TASC studies, the author divided them based on the strengths and weaknesses of these studies. In general, the scores received across all four studies ranged from 35.4% to 68% with an average of 55.2%. The TASC studies had the lowest average ratings across the models studied. The scores received for each quality indicator, its sub-indicators, and the total score for each study of TASC are presented in Appendix C.

Strengths of studies. The researchers in the four studies received high scores in two indicators: research conceptualization and measurements. The average score on the research conceptualization indicator was 4.2 out of five. The researchers of seventy-five percent of studies stated clearly the rationale for their research, while West (2008) provided no reasons for conducting her research. The purposes of the studies were written clearly.

For the measurements category, the researchers gave sufficient information about the instruments they used in their studies: an average of 4.45 was given for each sub-indicator. The researchers used more than one instrument to gather data. For example, West (2008) used a questionnaire, discussions with teachers and children, samples of children's writing, and observations to collect her data. Another strong sub-indicator in the measurement category was that the researchers provided sufficient descriptions of the ways they gathered data. For instance, Abu Awwad et al. (2014) used pretests and posttests to collect their data during the intervention.

Weaknesses of studies. Four indicators had the lowest scores. The researchers, in general, provided limited or no descriptions of the following: (a) settings and participants, (b) intervention, (c) fidelity of implementation (FOI), and (d) data analysis and outcomes. The average scores earned across these four indicators ranged from 1.17 to 2.18.

For the settings and participants category, the researchers in three studies did not provide any clear description of the methods they used to select students. This failure to provide information about the procedure for choosing participants makes replicating these studies in the future difficult. The only exception was the study by Faulkner (2008), in which students were a convenience sample, identified as talented in mathematics, and among the top ten percent of students in the UK. Another common weakness of this category was that discussions of the settings and resources were absent.

Intervention was similar to the settings and participants category, both earned a score of two out of five. In three studies, the researchers did not report any information about the teachers (e.g., how they implemented TASC in classrooms). Only Faulkner (2008) discussed TASC activities used by the teachers in classrooms. The fidelity of implementation was not assessed in any study, which may affect the validity of the results of the research.

The final weakness of TASC studies was data analysis and outcomes. The average score received by studies was 2.18. In studies by Davies (2008) and West (2008), the researchers did not illustrate how they dealt with the data they gathered. Neither of the two researchers provided any description of the method for analyzing data or of how each of them used trustworthiness and credibility to increase the confidence in their findings.

The Effect of the Use of TASC on Students' Creativity and Attitudes toward Learning

In the four studies selected for this review, the common purpose was to investigate the effect of the use of TASC on students' creativity and attitudes toward learning. The researchers in three studies examined the influence of the implementation of TASC on students' creativity in science, mathematics, and thinking skills (Davies, 2008; Faulkner, 2008; West, 2008), while the purpose of the fourth study (Abu Awwad et al., 2014) was to examine the influence of the use of the TASC wheel on students' self-directed learning readiness and academic self-efficacy.

The researchers concluded that the TASC wheel was an effective model in that it contributed to improving students' creativity in science, mathematics, and thinking skills. For example, one researcher investigated the effects of using TASC on students' thinking and problem-solving skills in science. She found in her qualitative study that after implementing TASC for six weeks, girls showed greater improvements in their creativity and thinking skills than did boys, while boys followed the TASC wheel's eight steps better than did the girls. However, the sample size of this study was small: only three females and three males participated (Davies, 2008).

In another study, the researchers found that the TASC wheel was a useful model that could be implemented in classrooms to improve attitudes toward learning. Abu Awwad et al. (2014) conducted their study with 73 girl students in grade 7 in one middle school. They used self-directed learning readiness and self-efficacy scales before and after the intervention to measure the outcomes of their study. The researchers found that students in the treatment group showed increases in self-directed learning readiness when compared to the control group. However, no differences were found between the treatment and control groups on academic self-efficacy.

Despite the promising results regarding the positive impact of TASC on students' creativity and attitudes toward learning, 50% of the studies (Davies, 2008; West, 2008) lacked essential research elements. The most important research elements the researchers did not illustrate clearly were the process they followed to nominate students to participate in their studies and how TASC was implemented in classrooms. In both studies, the researchers did not assess FOI to increase confidence in their findings. A summary of each study is presented in Table 7.

Table 7: *A Summary of TASC Studies*

Author(s)	Participants	Setting	Research Design	Purpose
Abu Awwad et al., 2014	• 73 students in grade 7	• One middle school	• Quasi-experimental	• Examined the influence of the use of the TASC wheel on students' self-directed learning readiness and academic self-efficacy.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Self-directed learning readiness scale • Self-efficacy scale	• ANCOVA	• Unspecified	• Self-directed learning readiness was increased among students in the treatment group than those of students in the control group. However, no differences were found between the treatment and control groups on students' academic self-efficacy.
Author(s)	**Participants**	**Setting**	**Research Design**	**Purpose**
Davies, 2008	• Six students in grade 6	• Elementary school	• Qualitative-action research	• Investigated the effects of using TASC on students' thinking and problem-solving skills in science.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Observations • Questionnaire	• Not stated	• Science	• Girls showed more improvement in their creativity and thinking skills than did the boys. However, boys followed the TASC wheel's eight steps better than did the girls.
Author(s)	**Participants**	**Setting**	**Research Design**	**Purpose**
Faulkner, 2008	• 35 talented mathematics students	• One private elementary school	• Qualitative-case study	• Examined the effect of the use of the TASC wheel to increase mathematical creativity among talented students.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Questionnaire • Group interview	• Not stated	• Mathematics	• Students believed that TASC was a good model and enjoyed solving mathematical problems in creative ways using TASC.

Table 7: *A Summary of TASC Studies*

Author(s)	Participants	Setting	Research Design	Purpose
West, 2008	• The number was not reported.	• Three elementary schools	• Mixed-methods	• Investigated the effect of implementing TASC on students' thinking skills and motivation toward learning.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Questionnaires • Discussions • Samples of children's writing • Observations	• Not stated	• Unspecified	• The researcher concluded that students became more independent learners and had a positive experience working with peers in groups to solve real-world problems. Their academic achievement, writing abilities, and motivation toward learning increased.

The Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses (DISCOVER) Curriculum Model

In this section, a summary of studies of the DISCOVER curriculum model was presented. The results of evaluation of the quality of DISCOVER studies and the effect of the use of DISCOVER on students' creativity in science, mathematics, and general creativity were reported.

A Summary of the DISCOVER Curriculum Studies

Six studies of the DISCOVER curriculum were found for this review. The researchers used quantitative designs to conduct their studies. Five studies were implemented in the USA in elementary schools (Jo & Maker, 2011; Maker et al., 2008; Maker, Rogers, Nielson, & Bauerle, 1996; Sak & Maker, 2006) or in both kindergarten and elementary schools (Maker et al., 2006), while the researchers in one study conducted their research in Taiwan in a preschool (Kuo, Su, & Maker, 2011). In four studies, no information was provided about the schools' areas while two studies were implemented in urban and rural areas (Jo & Maker, 2011; Kuo et al., 2011).

The classroom size in one study was comprised of less than 20 students, one study was implemented in classrooms that included between 20 and 29 students, while the classroom size in four studies was not reported. One study was conducted with gifted students, three studies were implemented in regular classrooms, while the researchers in two studies did not state the classroom type. All teachers who participated in all studies attended development professional workshops to implement the DISCOVER curriculum in their classrooms. No information was provided about teaching experience or teachers' genders in any of the studies. The subjects taught in these studies were science and mathematics, and the number of instruments for collecting data ranged from one to five, with an average of three instruments among the six studies. The common methods of gathering data were interviews, observations, and creativity tests. Appendix A has a descriptive classification of these studies of the DISCOVER curriculum.

The Results of the Evaluation of Quality of the DISCOVER Curriculum Studies

In general, the evaluation scores that DISCOVER studies earned were the highest among the four models in this review. The study by Maker et al. (2008) scored highest, meeting almost all quality indicators and receiving a total score of 95%, while the study by Kuo et al. (2011) scored lowest and earned only 83.6%. The other four studies had percentages ranging from 84% to 93.3%. The average percent across all six studies was 88.6%. The scores received for each quality indicator, its sub-indicators, and the total score for each study of DISCOVER are presented in Appendix D.

All studies received full scores on all sub-indicators of the research conceptualization category. The researchers of these studies stated clearly the rationale for their research, purpose of their studies, theoretical framework, and research questions. For example, Jo and Maker (2011) emphasized the importance of developing creativity in mathematics among students and the needs for assessing the levels of implementation of DISCOVER in classrooms. They supported their arguments with empirical evidence.

Similar to the research conceptualization, the researchers in all studies had high scores in the settings and participants category. The average score received was 4.27 out of 5. The researchers described in sufficient detail the settings, participants, and processes of selecting students to participate in the studies. In three studies (Maker et al., 2008; Maker et al., 2006; Sak & Maker, 2006), the description of the researchers' role in relation to the settings and participants was not stated.

For the intervention category, the average score earned was 4.3 out of 5. The interventionist's roles were described in each study, and the researchers provided information about how the DISCOVER curriculum was implemented in classrooms. However, the researchers in three studies did not provide any relevant demographic information about the teachers who implemented the DISCOVER curriculum (Jo & Maker, 2011; Kuo et al., 2011; Maker et al., 1996).

One of the unique aspects among all studies of the DISCOVER curriculum was assessing the fidelity of implementation (FOI) of this model. In all studies, full scores were given for each of the three sub-indicators: FOI, key features, and inter-observer reliability. Experts in the DISCOVER curriculum visited classrooms and gathered data on how teachers implemented DISCOVER, and then they interviewed the teachers to discuss their perceptions of this model.

Scores in the other two categories (measurements and data analysis) had an average of 3.94 and 4.67 respectively. The researchers in five studies used three to five instruments in their research, except for Sak and Maker (2006) who used only one instrument for collecting their data. All researchers described in detail their methods of gathering data. Only the researchers in two studies (Jo & Maker, 2011; Maker et al., 1996) did not discuss the validity of instruments they used in their research. However, Maker et al. stated that the DISCOVER instrument was new and that its validity and reliability needed to be supported by scientific evidence. The common data analysis techniques the researchers used were Multivariate Analysis of Variance (MANOVA) and Analysis of Variance (ANOVA). The effect sizes of all studies except Kuo et al. (2011) was reported.

The Effect of the Use of the DISCOVER Curriculum on Students' Creativity

The common purpose of the six DISCOVER studies was that the researchers examined the influence of the use of the DISCOVER curriculum on students' creativity. However, the researchers in all studies had different goals for their investigations. Across these studies, students benefited from DISCOVER as an intervention to promote their creativity in science, mathematics, and general creativity.

In four studies (Jo & Maker, 2011; Maker et al., 2008; Maker et al., 2006; Maker et al., 1996), the researchers examined the impact of the levels of implementation by the teachers on students' creativity. The studies were implemented in kindergarten and elementary schools, and the duration of these studies ranged from one to three years. However, the results of these studies varied based on grade and levels of implementation. For example, Jo and Maker (2011) explored the influence of the teachers' implementation of the DISCOVER curriculum on students' knowledge and creativity in mathematics. The DISCOVER curriculum was implemented for one year with students in grades 1 through 5. The researchers used a standardized test, a creativity test, observation, and interviews to gather their data. The results were statistically significant for students in grades 2 and 3. Students with high implementers in grade 2 had higher scores in creativity in mathematics compared to students with middle implementer teachers. All results in grade 3 were statistically significant among students in classrooms with high, middle, and low implementers in both creativity in mathematics and content knowledge. For students in grades 1, 4, and 5, the results were not statistically significant in both creativity in mathematics and content knowledge across the three levels of implementation. In contrast to Jo and Maker, Maker et al. (1996) investigated the impact of the teachers' implementation of the DISCOVER curriculum (high and middle) on students' problem-solving skills in spatial, logical-mathematical, and linguistic abilities. The researchers found a relationship between teachers' level of implementation and positive changes in problem-solving favoring students in high implementer classrooms.

Furthermore, Maker et al. (2008) investigated the association between general creativity and teachers' levels of implementation of the DISCOVER curriculum model during three years in a K-6 setting. The researchers concluded that creativity increased in general based on the levels of implementation. However, creativity was affected by the duration of implementation. For example, creativity increased among students in high or middle implementer classrooms during Years 1 and 2, creativity was stable among students in classrooms when the implementation level of their teachers increased from Year 1 to 2, and creativity decreased among students in classrooms when the implementation level of their teachers decreased from Year 1 to 2. However, the results in Year 3 were not statistically significant among all students related to differences in levels of implementation. A possible reason that students' creativity did not increase in Year 3 was that these students took the same creativity test, The Test for Creative Thinking-Drawing Production (TCT-DP), at the end of each school year and thus they became familiar with the tasks, so they likely were not taking the test very seriously. Similar to Maker et al. (2008), Maker et al. (2006) found that the differences in general creativity of students in Year 3 in classrooms at three levels of implementation were not statistically significant.

The DISCOVER curriculum was integrated with Multiple Intelligences (MI) and then compared to another teaching model, Talent Development, to examine the impact of the use of these two teaching approaches on gifted preschoolers' problem-solving abilities based on problem types. The researchers concluded that students performed high on closed and open problem-solving types, especially on Type 1 (closed), Type 4, and Type 5 (open) in both types of programs. However, students performed better on Type 4 than on Type 5 and better on both open-ended types when they were in the Talent Development program than when they were in the DISCOVER/MI program (Kuo et al., 2011).

Finally, the DISCOVER curriculum was used as an intervention to investigate the relationship between age, grade level, and domain-specific knowledge in the increase of students' creativity in mathematics as measured by originality, flexibility, elaboration, and fluency on the DISCOVER assessment. The researchers implemented their study in four elementary schools, from grade 1 to 5, in one year, using the DISCOVER assessments to gather their data. The researchers found that domain-specific knowledge was progressively associated with originality, flexibility, elaboration, and fluency from lower to upper grades. On the other hand, age was associated with creativity only in lower grades (Sak & Maker, 2006). A Summary of each study of DISCOVER is presented in Table 8.

Table 8: *A Summary of DISCOVER Studies*

Author(s)	Participants	Setting	Research Design	Purpose
Jo & Maker, 2011	• 835 students • 51 teachers	• Four elementary schools	• Quasi-experimental	• Explored the influence of the implementation of the DISCOVER curriculum on students' knowledge and creativity in mathematics.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Standardized test • Creativity test • Observations • Interviews	• MANOVA	• Mathematics	• The results were statistically significant for students in grades 2 and 3. Students with high implementers in grade 2 had higher scores in creativity in mathematics than students with middle implementers. All differences for grade 3 were statistically significant in creativity in mathematics and knowledge among students in classrooms with high, middle, and low implementers. However, the differences were not statistically significant for students in grades 1, 4, and 5.
Author(s)	**Participants**	**Setting**	**Research Design**	**Purpose**
Kuo et al., 2011	• 61 gifted preschoolers	• One preschool	• Quasi-experimental	• Examined the impact of the use of two teaching approaches (DISCOVER/MI and Talent Development) on gifted preschoolers' problem-solving abilities based on problem types.

(Continued)

Table 8: *A Summary of DISCOVER Studies*

	Data Source	Data Analysis	Subject(s)	Findings
	• Checklists • Interviews • Observation • Portfolio • Group intelligence tests • Individual intelligence tests	• ANOVA	• Science • Mathematics	• Students performed high on closed and open problem types, especially on Type 1 (closed), Type 4, and Type 5 (open) in both programs (DISCOVER/MI and Talent Development). However, students in the Talent Development program performed better on Type 4 problems than on Type 5, and better on both open-ended types than students did in the DISCOVER/MI program.
Author(s)	**Participants**	**Setting**	**Research Design**	**Purpose**
Maker et al., 2008	• 1986 Students, grade K to 6 • 108 teachers	• Four elementary schools	• Mixed method	• Examined the association between general creativity and teachers' levels of implementation of the DISCOVER curriculum model.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• The TCT-DP • Observations • Interviews	• ANOVA • ANCOVA	• Science • Mathematics	• Creativity increased among students in classrooms with high or middle implementers during Years 1 and 2, creativity was stable among students in classrooms when the implementation level of their teachers increased from Year 1 to Year 2, and creativity decreased among students in classrooms when the implementation level of their teachers decreased from Year 1 to Year 2.
Author(s)	**Participants**	**Setting**	**Research Design**	**Purpose**
Maker et al., 2006	• 2983 Students K-6 • 104 teachers	• Four elementary schools	• Quasi-experimental	• Examined the influence of the use of DISCOVER on students' general creativity.

(Continued)

Table 8: *A Summary of DISCOVER Studies*

	Data Source	Data Analysis	Subject(s)	Findings
	• The TCT-DP • Observations • Interviews	• ANOVA	• Unspecified	• Students in middle and high implementers' classrooms in Year 2 showed an increase in their general creativity. However, the differences in students' general creativity in Year 1 and 3 were not statistically significant across all classrooms regardless of the levels of implementation.
Author(s)	**Participants**	**Setting**	**Research Design**	**Purpose**
Maker et al., 1996	• 46 students from two classrooms	• Elementary school	• Quasi-experimental	• Investigated the impact of the levels of implementation of DISCOVER (high and middle) on students' problem-solving skills in spatial, logical-mathematical, and linguistic abilities.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• DISCOVER assessments	• T-tests	• Mathematics • Linguistics	• The researchers concluded that a significant relationship existed between teachers' levels of implementation and positive changes in students' problem-solving abilities.
Author(s)	**Participants**	**Setting**	**Research Design**	**Purpose**
Sak & Maker, 2006	• 841 students from grade 1 to 5	• Four elementary schools	• Quasi-experimental	• Examined the relationship among age, grade level, and domain-specific knowledge in the increase of students' creativity in mathematics as measured by originality, flexibility, elaboration, and fluency in the DISCOVER assessment.

(Continued)

Table 8: *A Summary of DISCOVER Studies*

	Data Source	Data Analysis	Subject(s)	Findings
	• The DISCOVER assessment	• Hierarchical regression analysis • MANOVA	• Mathematics	• Domain-specific knowledge was progressively associated with originality, flexibility, elaboration, and fluency from lower to upper grades. Age was associated with creativity only in lower grades.

Real Engagement in Active Problem Solving (REAPS) Studies

This section is comprised of three parts. First, a summary of studies of REAPS is reported. The quality of studies conducted to implement REAPS in classrooms is discussed in the second part. Finally, the effects of the use of REAPS on students' creativity in science and general creativity are reported, and students' perceptions of REAPS are discussed.

A Summary of REAPS Studies

Eight studies of REAPS were found for this review. The researchers in three studies used a quantitative method and implemented their research in three different countries: Australia (Alhusaini, 2016), Indonesia (Yulindar et al., 2018), and the USA (Zimmerman et al., 2011). The researchers in the other five studies used a qualitative method and conducted their research in the USA (Gomez-Arizaga et al., 2016; Reinoso, 2011), Australia (Wu et al., 2015), and New Zealand (Riley et al., 2017; Webber et al., 2018). Five studies were implemented in elementary schools and three studies were done in a high school (Riley et al.; Webber et al.; Yulindar et al.). Six studies were conducted in urban areas and two were implemented in rural areas (Riley et al; Webber et al.).

The classroom size in two studies was 20 to 29 students, one study was implemented in a classroom with more than 30 students, and researchers in five studies did not state the classroom size. Two studies were conducted with gifted students, two studies were implemented in regular classrooms, and the researchers in four studies did not report the classroom type.

Teachers in five studies (Alhusaini, 2016; Riley et al., 2017; Webber et al., 2018; Wu et al., 2015; Zimmerman et al., 2011) attended professional development workshops to implement the REAPS model in their classrooms, while the researchers in the other three studies did not state whether teachers received any professional development workshops. No information was provided about teaching experience or teachers' genders in all studies except Zimmerman et al. (2011), in which all teachers were females and had teaching experience of more than ten years. The subject taught in these studies was science. The number of instruments for collecting data ranged from one to four. Appendix A has a descriptive classification of these studies of the REAPS model.

The Results of the Evaluation of Quality of REAPS Studies

The scores received in the REAPS studies ranged from 48.2% (Yulindar et al., 2018) to 95% (Alhusaini, 2016). The other six studies had scores ranging from 55.3% to 92.7%. The average score across all eight studies was 73.7%. The scores received for each quality indicator, its sub-indicators, and the total score for each study of the REAPS model are presented in Appendix E.

Quantitative studies. The researchers in all studies received full scores on three sub-indicators of the research conceptualization. Each researcher stated clearly the rationale for the research, purpose of the study, and its theoretical framework. However, the researchers in two studies (Yulindar et al., 2018; Zimmerman et al., 2011) did not state research questions. The average score received on this indicator was 4.33 out of 5.

The measurements indicator had the highest average score (4.43 out of 5) among all studies. The researchers described in detail the methods of collecting data, and they used pretests and posttests for their studies. Researchers in only one study (Yulindar et al., 2018) used one instrument to gather their data. However, they did not discuss the validity and reliability of the test they used, the Test of Problem Solving Ability. The absence of reporting the validity and reliability in any instrument affected the credibility of the results of the study.

For three indicators, setting and participants, fidelity of implementation (FOI), and data analysis, the average scores received were 3.3, 2.5, and 3.8 respectively. One study (Yulindar et al., 2018) negatively affected the average scores on these indicators. For example, these researchers did not provide any information about settings, participants, or the researchers' role in relation to the settings and participants. Also, the researchers did not assess the FOI, which might influence the findings.

Qualitative studies. In general, the researchers who used a qualitative design had higher average scores in the quality of their studies (74.3%) than the researchers in quantitative studies (72.7%). Wu et al. (2015) scored the highest (92.7%), while Reinoso (2011) scored the lowest (55.3%).

On five indicators, the average score received ranged from 4 to 5. The researchers in two studies (Gomez-Arizaga et al., 2016; Wu et al., 2015) provided sufficient information about the rationale that led them to conduct their research, the theoretical framework, settings and participants, and collection and analysis of data; the researchers also reported their findings in sufficient detail. However, the remaining sixth indicator, intervention, had a low score (3 out of 5). The researchers in all studies did not describe the role of teachers in classrooms with exception for Webber et al. (2018) and Riley et al. (2017) who provided information of the role of teachers in classrooms. No sufficient demographic information about teachers was provided in all studies. Researchers in three studies (Riley et al., 2017; Reinoso, 2011; Webber et al., 2018) discussed in detail how the REAPS model was implemented in the classroom.

The Effect of the Use of the REAPS Model on Students' Creativity and Motivation toward Learning

Across all the seven studies, the researchers found that REAPS was an effective model for increasing creativity and improving attitudes toward learning (Alhusaini, 2016; Gomez-Arizaga et al., 2016; Riley et al., 2017; Reinoso, 2011; Wu et al., 2015; Yulindar et al., 2018; Zimmerman et al., 2011). This model can be implemented in different school settings such as elementary and high schools.

The REAPS model was designed to increase students' creativity, especially if the students were exposed to this model for a long period of time. For example, the REAPS model was implemented in an elementary school with 360 students from grade 1 to 6 to investigate the differences in students' general creativity and creativity in science after exposure to REAPS in two different durations, long and short. The differences between the two groups on a posttest of general creativity were not statistically significant. However, students who were exposed to the REAPS model for a long duration showed greater increases in their scientific creativity than students exposed for a short duration (Alhusaini, 2016).

Students' problem-solving ability improved after REAPS was used as an intervention. For example, one study was implemented in a high school with 35 students to investigate the effect of the use of the REAPS model on students' problem-solving abilities in a teaching unit on the concept of heat transfer. The researchers used pretests and posttests to gather their data using the Test of Problem Solving Ability. They found that students' problem-solving ability increased after implementing the REAPS model in a physics classroom.

Not only did REAPS have a positive impact on students' creativity and problem-solving skills but it also played a crucial role in increasing motivation toward learning (Gomez-Arizaga et al., 2016; Riley et al., 2017; Reinoso, 2011; Webber et al., 2018; Wu et al., 2015). For instance, 42 students in one elementary school were interviewed after they had participated in classrooms whose teachers used REAPS as an intervention. The students expressed positive attitudes toward REAPS, and they believed it was an effective teaching model. They emphasized that this model could contribute to improving their academic abilities and skills (Wu et al.).

Even though the researchers in all seven REAPS studies concluded that the REAPS model had positive effects on students' creativity and motivation toward learning, the researchers in two studies (Reinoso, 2011; Yulindar et al., 2018) received scores on the quality indicators (CEC, 2008) of less than 60%, which raises questions about the credibility of their findings. In both studies, the researchers did not explain the procedure for selecting students, how REAPS was implemented in classrooms, or the process for analyzing the data, all of which decreased confidence in their results. A summary of each study of REAPS is presented in Table 9.

Table 9: *A Summary of REAPS Studies*

Author(s)	Participants	Settings	Research Design	Purpose
Alhusaini, 2016	• 360 students from grade 1 to grade 6	• One elementary school	• Quasi-experimental	• Investigated the differences in students' general creativity and creativity in science after exposing them to the REAPS model in two different durations. Also, the researcher aimed to explore which aspects of creative problem solving were most affected by the long duration of the intervention.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• TCT-DP • TCPS-S	• ANCOVA • Binary logistic regression	• Science	• No statistically significant differences were found between the two groups of students (long and short duration) in the posttest scores on the Test for Creative Thinking Drawing Production (TCT-DP). • Statistically significant differences were found between the two groups of students on the posttest scores on the Test of Creative Problem Solving in Science TCPS-S, favoring the long duration group. • Generating ideas, adding details to ideas, and finding problems were the aspects most affected by the long duration of the intervention.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Gomez-Arizaga et al., 2016	• 24 students in grade 3	• One elementary school	• Qualitative study	• Investigated students' perceptions after exposure to the Full Option Science System (FOSS) and the REAPS model in a science classroom.

(Continued)

Table 9: *A Summary of REAPS Studies*

Author(s)	Data Source	Data Analysis	Subject(s)	Findings
	• Interview • Drawing	• Coding, categorizing, and developing themes.	• Science	• Students had opportunities to share, create, and put their ideas into action after the REAPS model was implemented in their classroom.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Reinoso, 2011	• 24 students from grade	• One elementary school	• Qualitative - action research	• Examined the effect of the use of the REAPS model on students' creative problem solving.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Students' products	• Not stated	• Science	• Students expressed enthusiasm and involvement in learning. They developed their creative problem solving skills to deal with real problems.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Riley et al., 2017	• 90 students and three teachers	• One High school- grade 9	• Qualitative- case study	• Determined the adaptability and effectiveness of REAPS in increasing engagement and achievement among students in one rural high school.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Interviews • Observations • Documents • Surveys	•Not stated	•Science	• Student showed greater engagement and collaboration. The REAPS model contributed to identify hidden talents and some shifts in achievement.

(Continued)

Table 9: *A Summary of REAPS Studies*

Author(s)	Participants	Settings	Research Design	Purpose
Webber et al., 2018	• Students in grades 8-9 and ten teachers	• Two High schools	• Qualitative	• Determined the adaptability and effectiveness of REAPS in increasing engagement and achievement among gifted students in two rural high schools
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Interviews	• Not clear	• Science	• The adaptation of REAPS was effective in engaging and increasing the success of gifted and talented Maori boys.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Wu et al., 2015	• 42 students	• One elementary school	• Qualitative	• Investigated participants' perceptions of the REAPS model.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Interviews • Artifacts	• Coding, categorizing, and developing themes.	• Science	• Students expressed positive attitudes toward REAPS, and they believed it was an effective teaching model that helped them to improve their academic abilities and skills.
Author(s)	**Participants**	**Settings**	**Research Design**	**Purpose**
Yulindar et al., 2018	• 35 students	• One high school	• Quasi-experimental	• Examined the effect of the use of the REAPS model on students' problem-solving abilities in a teaching unit on the concept of heat transfer.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• The test of problem-solving Ability	• Mean scores	• Science	• Students' problem-solving ability increased after the REAPS model was implemented in a physics classroom.

(Continued)

Table 9: *A Summary of REAPS Studies*

Author(s)	Participants	Settings	Research Design	Purpose
Zimmerman et al., 2011	• Two classrooms in grade 3. No information about students' number	• One elementary school	• Quasi-experimental	• Examined the effect of using concept maps as a tool to solve real problems in earth science after combining the REAPS model with the Full Option Science System.
	Data Source	**Data Analysis**	**Subject(s)**	**Findings**
	• Concept maps	• T-test	• Science	• Students' knowledge increased in general and REAPS contributed to improving the teacher's skills to teach the FOSS unit.

Summary of Chapter II

Forty-four studies were included in the literature review: 26 PBL studies, four on TASC, six on DISCOVER, and eight on REAPS. The purpose of this chapter was to answer two questions: (a) to what extent did studies on PBL, TASC, DISCOVER, and REAPS meet the quality indicators that were modified from the Council for Exceptional Children (2008)? and (b) to what extent did students' general creativity, creativity in science and mathematics, and academic achievement increase after they were exposed to the PBL, TASC, DISCOVER, and REAPS models?

To answer the first question, quality indicators developed by the Council for Exceptional Children (CEC) in 2008 were used to assess the quality of quantitative, qualitative, and mixed-methods studies. For the 30 quantitative studies, the author used seven quality indicators. Scores ranged from 36% to 95.9%; the DISCOVER studies ranged from 83.6% to 95% and had the highest average score. For the 10 qualitative studies, scores ranged from 49.1% to 92.7%; the five REAPS studies had the highest average score (74.3%). A mixed-methods design was used in four studies (three PBL and one TASC), and scores ranged from 35.4% to 84.8%. The highest and lowest scores received on quality indicators based on research design are shown in Appendix F.

The results across all studies of PBL, TASC, DISCOVR, and REAPS were promising. The researchers of the 44 studies concluded that these four models were effective either when they were used alone as interventions or combined with other teaching strategies such as cooperative learning. However, researchers in ten studies (six PBL, two TASC, and two REAPS) received scores lower than 60% when their research was evaluated using the quality indicators, which raises concerns about the credibility of the findings of these studies.

Students who were exposed to PBL, TASC, DISCOVR, and REAPS benefited from these models, and their creativity in general and academic achievement increased. Students emphasized that their attitudes toward learning and motivation changed positively after these teaching models were implemented in their classrooms.

CHAPTER III: METHOD

About the Study

This study was part of a cooperative project started in Fall 2013 between the University of Arizona in the United States and one public elementary school in Australia. The principal investigator of this project and her team collaborated with the school principal to implement the Real Engagement in Active Problem Solving (REAPS) model. The research team consisted of a professor who an expert in education of the gifted, a scientist, and a teacher with over 30 years of experience teaching and supervising teachers, especially at the elementary level. The aim of this project was to develop students' ability to solve real problems in creative and effective ways. Before implementing the REAPS model, the research team gathered parental consent forms (Appendix G). They then provided one professional workshop for five days to introduce the REAPS model to teachers and help them differentiate the curricula and meet their students' needs. The teachers received follow-up professional support from the research team and administrators in the school. In the follow-up professional support, the research team visited the school two times per year and also reviewed teaching units that were developed by the teachers based on the REAPS model. After each observation, the research team member who observed the teacher met with the teacher to discuss the observation and make suggestions for the future. They provided written feedback for each teacher about their observations and reviews of their teaching units. The research team also collaborated with the school principal by email to answer any questions or help with challenges teachers faced. In this study, I used the data that were gathered in 2015 (pretests) and 2016 (posttests).

Research Design

The main purpose of this study was to examine the influence of levels of Fidelity of Implementation (FOI) on student creative problem solving in science. Because the REAPS model was implemented school-wide with no treatment and control groups and the participants were not selected randomly, the best approach for conducting this study was to use a quasi-experimental design (Abbott, 2017). In experimental and quasi-experimental designs, the main purpose is to test the impact of an intervention and its outcomes (Creswell, 2009). Five dependent variables were used to answer Questions 1 and 2: the total scores on TCPS-S, fluency, flexibility, elaboration, and originality. Grade levels were used as an independent variable for Question 1 and levels of FOI were used as an independent variable for Question 2. For Question 3, three dependent variables were used (Finding Problems, Generating Detailed Solutions, and Classifying Elements) with one independent variable (levels of FOI).

Setting and Participants

Setting

The study was conducted in one urban public elementary school (K-6) in New South Wales, Australia. Students came from varied ethnic groups, backgrounds, and economic levels. Approximately 52% of the students were from language backgrounds other than English, with more than 50 languages represented. In the school, a number of research-based education programs were offered to meet students' learning needs, such as acquisition of English as a second language (ESL) and accommodations of learning difficulties. Teachers designed their teaching units based on common themes such as change and functionality, culture and identity, and design and systems.

The school year was based on a quarter system with four terms. The duration of each term was ten weeks, and students had a two-week holiday at the end of each term. The average classroom size was 25 students. In Year 1, the total number of classrooms from grades 2 to 5 was 17: five classes in grade 2; four classrooms in each of grades 3, 4, and 5. In Year 2, the total number of classrooms from grade 3 to 6 was 16: five classes in grade 3; four classrooms in each of grades 4 and 5; and three classes in grade 6. Classrooms contained tables of different sizes for individual and group work. Ample equipment, technology devices, and educational resources were provided in each classroom, such as a computer, a projector, speakers, access to the Internet, art supplies, manipulatives, and books.

Participants

The total number of participants in Year 1 was 317: 88 second grade students (27.8 %), 79 third grade students (24.9%), 77 fourth grade students (24.3%), and 73 fifth grade students (23%). In Year 2, the total number of participants was 317: 88 third grade students (27.8 %), 79 fourth grade students (24.9%), 77 fifth grade students (24.3%), and 73 sixth grade students (23%). Student information such as gender, socioeconomic status, and ethnicity was not reported in this study due to the school policy. Each parent received a letter from the principal investigator and the school principal (Appendix H) to discuss the rights of their children who participated in this study. Students could withdraw from the study at any time, abstain from any pretest or posttest, and be assured of the protection of privacy. Students who completed only the pretest or posttest were excluded from this study because it was not possible to measure the change in scores if one of the tests was not taken. The number of students who were excluded is shown in Table 10. Students who were in a specific classroom in 2015 were mixed with other students in different classrooms in 2016. For example, 20 students who were in grade 2 in classroom AB (pseudonym) were mixed with other students in different classrooms in grade 3 in 2016 (nine students moved to classroom AC, 6 students to classroom AD, four students to classroom AE, and one student to classroom AF).

Table 10: *Number of Students Excluded from This Study*

Grade Level	Pretest only	Posttest only	Total by Grade Level
Grade 2Y1 & 3Y2	16	13	**29**
Grade 3Y1 & 4Y2	15	13	**28**
Grade 4Y1 & 5Y2	9	5	**14**
Grade 5Y1 & 6Y2	7	2	**9**
Total	**47**	**33**	**80**

Note: Grade 2Y1 & 3Y2 = Students who were in grade 2 in Year 1 and then entered grade 3 in Year 2; Grade 3Y1 & 4Y2 = Students who were in grade 3 in Year 1 and then entered grade 4 in Year 2; Grade 4Y1 & 5Y2 = Students who were in grade 4 in Year 1 and then entered grade 5 in Year 2; Grade 5Y1 & 6Y2 = Students who were in grade 5 in Year 1 and then entered in grade 6 in Year 2

Intervention

About the Real Engagement in Active Problem Solving (REAPS) Model

The REAPS model is a combination of three teaching models: the Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses (DISCOVER) curriculum, Thinking Actively in a Social Context (TASC), and Problem-Based Learning (PBL). These models complement each other, and each one of them has its contribution in the REAPS model (Maker & Zimmerman, 2008). The philosophy of differentiation of instruction was integrated with the REAPS model to serve all students and to help teachers to design their teaching units to meet students' needs.

The purpose of differentiated instruction is to help students develop as individuals based on their readiness and learning styles to improve their abilities and skills (Maker, 1982; Tomlinson, 2001). Thus, the developers of REAPS used the four components of differentiated instruction in their model: *content, process, product,* and *learning environment*. Each component has its own differentiation principles.

Content is what to teach and it includes concepts, principles, and skills used in reaching the learning goals. Content includes six principles: *abstractness, complexity, variety, study of people*, *study of methods,* and *organization for learning value*. Process is the method by which teachers deliver the content to their students, such as integrating problem-solving strategies, and processes that students must use. Teachers should use eight principles to guide differentiation of processes: *higher levels of thinking, open-endedness, discovery, evidence of reasoning, freedom of choice, group interaction, pacing,* and *variety*. Products are any products students develop or create to show what they know and can do, such as written papers, drawings, and presentations. Six principles guide differentiation of products: *real problems, real audiences, transformation, variety, self-selected format,* and *appropriate evaluation*. The learning environment is the settings and environments that allow teachers to reach their learning goals. It includes eight principles: *learner-centered, independence, openness, acceptance, complexity, varied groupings, flexibility*, and *high mobility*.

A teacher who implements this model must differentiate these four components by first designing a teaching unit based on a real local problem that is age-appropriate for the students. However, the teacher needs to make adjustments in the classroom by providing scaffolding for struggling students, illustrating a good model for successful performance, and encouraging students to meet challenges. The teacher is a facilitator who leads students through a step-by-step problem-solving process. The teacher provides a flexible learning environment in which students can move freely inside and outside the classroom based on the problem they attempt to solve. The teacher provides many resources to the students and is open to new ideas.

During implementation of the REAPS model, students work to solve a problem from stakeholders' perspectives either in small groups, large groups, or individually. They gather new information, apply knowledge, develop new ideas and solutions, evaluate their solutions based on criteria that are developed by them and/or by their teacher, and communicate their ideas to other groups in their classrooms.

How the REAPS Model was Implemented

The REAPS model was implemented first in cluster classrooms in which gifted students were served. Before implementation, all teachers of cluster classes were observed and interviewed by two members of the research team prior to their attendance at a five-day professional development workshop. In the workshop, the teachers were introduced to the REAPS model and its components: the DISCOVER curriculum model, TASC, and PBL.

Teachers designed teaching units based on the conceptual framework of the school's curriculum and the framework of the REAPS model. The research team reviewed the teaching units and gave the teachers feedback. Prior to implementation, all students in the school were given a pretest to measure their general creativity, creative problem solving in science, creative problem solving in math, and understanding of the complexities and interrelationships among concepts. After the implementation of the REAPS model for the first quarter of the school year, the school's administrators, teachers, and parents decided to expand the use of REAPS to include all classrooms in the school. The REAPS model was then implemented during the second and fourth academic quarters.

Teachers who attended the REAPS workshop mentored other teachers in the school. For each grade level, teachers designed their REAPS lessons based on a case study from each grade level that was developed from a macro concept. For example, in grade 2, the macro concept in quarter 2 was "heritage and tradition." The teachers created a focus question around which to design their lessons for the quarter: The school wanted to create a memorial to commemorate an important local person from the past. Who should we remember and why? What memorial should we create? Figure 5 illustrates how the intervention was implemented.

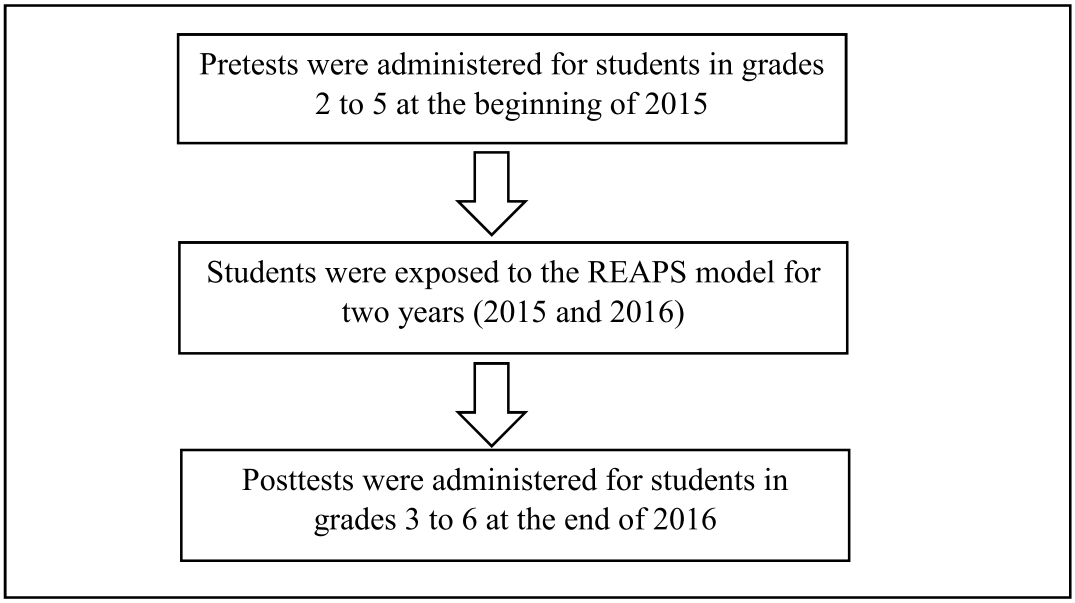

Figure 5. How the Students Participated in the Intervention in this Study.

An Example of a Teaching Unit

Content. To develop a teaching unit, teachers need to decide what to teach. They can develop content at different levels, starting from the lowest level (basic facts or phenomena), moving up to group these facts or phenomena (as concepts) based on their common characteristics, going up further to identify relationships between and among concepts (principles), proceeding to find relationships between and among principles within and across academic disciplines (theories), and reaching the highest level (macro concepts) by connecting interdisciplinary ideas that can be found in various academic topics such as change, systems, interdependence, sustainability, and patterns.

In this example, teachers defined interrelationships and sustainability as macro concepts that were found among different academic subjects in grade 4: science, history, and geography. They worked as a group to develop the teaching unit. They organized the facts and phenomena, concepts, principles, and theories around these macro concepts. They chose a real life problem related to tourist development on land considered sacred by the Aboriginal people, and created the following case study with three focus questions:

> A development company has proposed a new tourist resort with a range of facilities and recreational activities in Goowoola, a pristine, old-growth rainforest.
>
> 1. What will be the impact on the ecosystems of Goowoola Rainforest?
> 2. Should it be approved or denied?
> 3. Is there a compromise compatible with the long-term goal of protecting this rainforest?

The teachers incorporated all six principles for content differentiation into the teaching unit. The problem of the case study was based on a macro concept (abstractness): it was organized around interrelationships and sustainability (organization for learning value), the students were divided into different stakeholder groups to study the problem from different perspectives (complexity), content from different disciplines was integrated (variety), students were encouraged to study the problem based on stakeholders' perspectives (study of people), and they were encouraged to explore and examine similar real life examples to find a method to solve their problem (study of methods).

Process. Teachers incorporated process principles with the following TASC steps: *Gather/Organize*, *Identify*, *Generate*, *Decide*, *Implement*, *Evaluate*, *Communicate*, and *Learn from Experience* (Adams & Wallace, 1991; Wallace et al., 2012). Even though all process principles were incorporated in the teaching units, only some of them were observed in this example. Before students started solving the problem, the teacher pretested them to measure what they already knew about the problem to avoid teaching unnecessary content. Group interaction and learning from experience were integrated in all eight steps of TASC.

During the first step, Gather/Organize, the teachers introduced four stakeholder groups to the students: Tourism Centre Development Association, Indigenous Council, Residents Association, and Green Environmental Agency. The teachers provided several resources such as articles, reports, and books so that each group could collect and organize information to investigate the problem based on the perspective of their stakeholder, and students also gathered their own information about the problem. Some principles for process differentiation were incorporated at this step. For example, discovery was achieved when each student learned new information and made new connections.

In the second step of TASC, Identify*,* each group worked to design a problem that might arise from their stakeholder's perspective if the company were given permission to build a tourist resort. They collaborated to form a question that framed their perception of the problem based on the roles of their stakeholders in real life. For example, students in the Indigenous Council stakeholder group defined the problem that would affect this group if the resort were constructed. Evidence of reasoning and open-endedness principles for process differentiation were combined at this step. For instance, the teacher asked each group to explain the reasons that led them to select their problem.

In the next step, Generate, each group used brainstorming to produce as many ideas as they could to solve problem they defined. At this step, teachers reminded students of the rules of brainstorming such as avoiding judging a group member's ideas. Each group listed its ideas and then, if needed, gathered and organized new information that they might have missed. At this step, several differentiation process principles were integrated: higher levels of thinking, group interaction, open-endedness, pacing, and variety.

At the Decide step, each group developed a list of criteria to evaluate the ideas they generated at the previous step. Each idea was examined, and then the group selected the solution they thought was best to solve their problem. In some cases, teachers helped groups that had difficulties developing their own criteria. The freedom of choice principle was incorporated at this step because students had an opportunity to select their own criteria for evaluating the solutions. Higher levels of thinking (evaluation), group interaction, pacing, and variety were other principles used.

At the fifth step, Implement, groups worked to create a product that reflected their solution. The product could be tangible or intangible. For example, a 3D model might be developed to illustrate how the problem could be fixed, or students could make a PowerPoint presentation to describe their solutions. Some groups went back to the first step, Gather/Organize, to collect more information about their solutions. The principle of higher levels of thinking was integrated at this step because students had an opportunity to implement their own solutions.

In the next step of TASC, Evaluate, each group assessed the implementation of the product they developed. Some groups used the same criteria that were developed in the fourth step (Decide), and some groups developed new criteria to evaluate their products. Communication was the next step. Each group developed its own method to present its product to other groups in the classroom. They used artistic, dramatic, writing, and technology skills in their presentations. Teachers also invited guests such as school staff and parents to attend the groups' presentations.

At the final step of TASC, Learn from Experience, students explained what they had learned about the content and the problem solving process. They also evaluated their own participation in the group.

Product. All principles for product differentiation were used to guide students to develop their final products. Students worked to solve a real problem and this problem was not contrived (result from real problems). Each group evaluated all solutions using criteria they developed and then selected the best solution based on their criteria (appropriate evaluation). While groups were working to develop their products, the teacher encouraged them to transform information they gathered to develop original products rather than use existing products (transformation and variety). Each group had the freedom to select any formats that were appropriate for their products (self-selected format): charts, graphs, models, diagrams, dramatic performances, speeches, PowerPoint presentations, and movies. When groups completed their products, they presented them to real audiences made up of classmates, parents, and school staff.

Learning Environment. The teachers integrated all eight differentiation principles for learning environments. The teacher's role was that of a facilitator leading students through the TASC steps (learner-centered). They encouraged all students to be independent learners to find information that helped them to solve the problems. The teachers were open to new ideas, used varied processes, incorporated new information, and accepted different perspectives from the students (openness and acceptance).

Students had opportunities to go outside the classroom to gather more information, and in some cases some students went outside of school with the proper supervision to find additional resources and/or to collect data (high mobility and complexity). Teachers allowed students from different groups to work together to share their experiences with each other (flexibility and varied groupings). For example, students who were responsible for creating PowerPoint presentations in their respective groups worked together to learn and practice their PowerPoint skills.

Instruments

Five instruments were employed to gather the data. These instruments included the Test of Creative Problem Solving in Science (TCPS-S), classroom observations, interviews, self-report inventories, and evaluation of teachers by school curriculum supervisors.

The Test of Creative Problem Solving in Science (TCPS-S)

The TCPS-S was modified from an instrument designed by Mohamed (2006) to measure students' ability to solve problems creatively in science (Maker et al., 2017). The TCPS-S consisted of two tasks and could be completed within 45 minutes. Task 1, Problems and Solutions, contained three sub-tasks: (a) Identifying Problems (1A), (b) Generating Solutions (1B), and (c) Designing a Solution (1C). In Task 1A, students listed local environmental problems that could be seen in four pictures. In Task 1B, students generated solutions to one of the problems each student identified in Task 1A. In Task 1C, students selected one solution from the list in Task 1B and chose one of three options: (a) make a drawing, (b) write a description, or (c) make a drawing and write a description (Maker et al.). In Task 2, Classifying Elements, students grouped pictures of flowers based on similar characteristics and named each group. The construct validity of the TCPS-S was supported by evidence from factor analysis; three factors of the TCPS-S accounted for 63.96% of the overall variance in creative problem solving in science (Maker et al).

Classroom Observation

The research team observed teachers in their classrooms twice per year. Each visit included one or two observers and lasted at least one hour. During each observation, the research team took pictures of classrooms and students' products, wrote notes, and gathered relevant documents that were given to students. Later, each member of the team completed the observation form to help them organize their comments. The form contained five sections: differentiation of content, differentiation of process, differentiation of products, differentiation of learning environment, and general comments, with a list of behaviors to be observed for each section (Appendix I).

After each visit, if any observer had questions about the purpose of a learning experience the teacher used during the lesson, the teacher was asked to explain or write it on the observation form. After the observation form was completed, every teacher had an opportunity to review it and add comments. The research team also had group and individual discussions with teachers to discuss their understanding of the REAPS principles.

Interviews

The research team interviewed all teachers in this study to determine their perceptions of the REAPS model. Each interview lasted approximately 45 minutes. Interview questions included the following four examples: (a) If I say, "REAPS," what words come to your mind? (b) What do you think are the most valuable aspects of the REAPS model? Please explain why.

(c) What do you think are the most challenging aspects of implementing the REAPS model? Please explain why. (d) What, if anything, do you think needs to be added to the REAPS model to make it more effective?

Self-Report Inventories

Teachers were asked to complete a self-report inventory once during this study to reflect on and assess their own implementation of the REAPS model in their classrooms. They completed a checklist that was developed by the research team and contained four sections: content, process, product, and learning environment; the form also included a section for general comments. In each of the four main sections, a list of behaviors that belonged to the components of the REAPS model (DISCOVER, PBL, and TASC) was included. In this form, the teachers used a 6-point scale to evaluate their implementation of REAPS, with 0 = I do not know how to do this; 1 = I know how to do this, but I never do it; 2 = I seldom do this; 3 = I sometimes do this; 4 = I often do this; 5 = I always do this when I do REAPS; and 6 = This is an integral part of my teaching practice whether or not I am doing REAPS. Part of this form is presented in Appendix J.

Evaluation of Teachers by School Curriculum Supervisors

The research team asked two curriculum supervisors at the school to evaluate teachers and send their reports to the research team. The checklist for assessment of implementation of REAPS by supervisors is in Appendix K and it contained the same items of the teacher evaluation form (Self-Report Inventories).

Data Collection Procedures

The Test of Creative Problem Solving in Science (TCPS-S)

Pretests were administered in the second quarter of 2015 and posttests were implemented in the fourth quarter of 2016 to measure creative problem solving in science. For both the pretest and the posttest, the classroom teacher conducted an introductory activity with the whole group of students to demonstrate what to do when answering the questions on the TCPS-S. After practicing with an example, students completed Task 1 in which they observed three pictures of the local natural environment in Sydney and were asked to (a) identify problems from the pictures, (b) generate solutions, and (c) implement a solution. In Task 2, each student was given 18 picture cards of flowers and was asked to (a) make as many groups of flowers as he/she could, (b) give each group of flowers a name, and (c) write details about how the flowers in each group were alike.

Scoring the TCPS-S. Before scoring students' responses on the pretests, a database was developed to include all students' answers on the TCPS-S sub-tasks. One member of the research team first read all students' responses and entered them in Excel files. For example, in grade 4, all students' responses were read and entered in Excel files, one for each sub-task. Three data tables were then developed to sort the frequency of students' responses. The purpose of developing the database was to determine original responses for each grade level (Table 11). Then, two postdoctoral researchers and two doctoral students scored the pretests using the scoring procedure described by Maker and her colleagues (Maker et al., 2017). The scoring system for the TCPS-S is presented in Table 12.

To evaluate the performance of each student on Tasks 1A, 1B, and 2, fluency, flexibility, elaboration, and originality were scored (Guilford, 1950; Torrance, 1963; 1988). Fluency was

Table 11: *Example of Originality Scoring in 4th Grade*

Tasks	Responses	Originality Scores
Task 1A (Identifying problems)	Energy, Electricity, Greenhouse, Food Chain	5
	Ozone Layer, CO_2, Sunlight Blocking, Don't Care	3
	Storm, Tsunami, Global Warming, Whaling	1
	Rubbish, Trash, Boat, Broken Bridge, Pollution, Animals Die, Smoke, Oil Spill, No Tree, No Space	0
Task 1B (Generating Solutions)	Sustainable, Get Someone to Clean, Tell the Police, Jail	5
	Government, Watching People	3
	Make a Rule, Fine, Use Less Stuff	1
	More Bins, Bring Bag, Don't Litter, Keep Clean, Pick up the Rubbish, Clean the Air, More Plants, Stop Polluting, Recycle, Fix the Dock	0
Task 2 (Classifying Elements)	Peaceful, Exploded, Tropical, Smooth, Spines, Scaly, Bubbly, Nectar, Sponge, Skyscraper	5
	Eyes, Tentacles, Egg, Insect Food, Star, Cloud, Bird Food, Warriors	3
	Fiery, Find in House, Rainbow, Power, Circle, Texture, Dessert, Winter	1
	Colors, Flowers, Have Leaves, Number, Cute, Ugly, Popular, Odd, Fruit, Bright, Garden, Flying, Wild, Male, Female, Festival	0

Note: This table was introduced by Maker, C. J., Jo, S., Alfaiz, F. S., & Alhusaini, A. A. (2017). *The test of creative problem solving in science (TCPS-S): Construct and concurrent validity.* Manuscript in preparation. Department of Disability and Psychoeducational Studies, University of Arizona, Tucson, Arizona, USA.

Table 12: *Scoring System for the TCPS-S*

Item	Measured Component	Procedure	Point Range
Task 1 (a and b)	Fluency	How many scientifically valid responses were written?	1 point each
	Flexibility	How many categories were found in the scientifically valid responses?	3 points each
	Originality	How many responses were statistically infrequent and different from what other students listed?	5 (below 2%), 3 (2-5%), and 1 (6-10%)
Task 1 (c)	Creative Product	Using the CAT, experts judged the level of creativity of students' drawings.	1 to 7 levels
Task 2	Fluency	How many scientifically valid responses were written?	1 point each
	Flexibility	How many categories were found in the scientifically valid responses?	3 points each
	Originality	How many responses were statistically infrequent and different from what other students listed?	5 (below 2%), 3 (2-5%), and 1 (6-10%)
	Elaboration	How many meaningful details were added to the responses?	3 points each

Note: This table was introduced by Maker, C. J., Jo, S., Alfaiz, F. S., & Alhusaini, A. A. (2017). *The test of creative problem solving in science (TCPS-S): Construct and concurrent validity.* Manuscript in preparation. Department of Disability and Psychoeducational Studies, University of Arizona, Tucson, Arizona, USA.

determined by the number of responses that were considered scientifically valid, with one point given for each valid response. For example, "greenhouse," "global warming," and "pollution" were considered valid scientific responses. On the other hand, a description of an action, such as "mini boat is behind the other boat" and "boat cannot move back", were not considered valid scientific responses. The flexibility score was based on the number of categories of responses, with three points given for each valid category. Elaboration was defined as the number of details in each answer, with three points given for each detailed response that was scientifically valid. Originality was scored on a scale of 1 to 5 points based on the frequency of students' responses, with more points given for a higher level of originality (e.g., for each response that appeared with a frequency of less than 6 to 10%, 1 point was given; for a response with a frequency of less than 2%, 5 points were given; Maker et al., 2017). To check inter-rater reliability, the raters first scored one classroom at each grade level; the inter-rater reliability for Task 1A ranged from .77 to .84, that of Task 1B ranged from .85 to .91, that of Task 2 ranged from .83 to .93, and the overall inter-rater reliability among the tasks ranged from .87 to .94.

Three postdoctoral researchers and one doctoral student rated Task 1C using the Consensual Assessment Technique (CAT; Amabile, 1983; 1996). Before the raters assessed all students' responses, one classroom from each grade level was selected and scored by each rater to establish inter-rater reliability. The value of inter-rater reliability ranged from .86 to .93. The raters used a 7-point scale to score students' creative problem solving when developing a solution to an environmental problem; the average of the scores was used. For example, in grade 5, 7 points were given to a student response if he/she drew or wrote a solution that could protect the local environment and that was unique or rare, while one point was given for a student response if he/she drew or wrote a solution that was not unique, such as putting rubbish in a trash container. Most students used drawings to answer Task 1C. The raters might be influenced by the quality of students' drawings instead of the uniqueness of solutions (Amabile, 1996). Thus, each rater scored two classrooms according to the quality of students' drawings. The value of inter-rater reliability of drawings ranged from .76 to .90. Then the correlation was calculated between the score of the quality of drawings and the score of the uniqueness of solutions; the correlation ranged from .17 to .27. This correlation showed that students' drawing ability did not influence raters' scoring of the creativity of their responses.

The author and one of his colleagues scored the posttests following the same procedures that were implemented in scoring the pretests. First, they scored one classroom for each grade level; the inter-rater reliability for Task 1A was .91, that of Task 1B was .87, that of Task 1C (the uniqueness of solutions only) was .81, that of Task 1C (the uniqueness of solutions and quality of drawings) ranged from .15 to .25. The low correlation between the uniqueness of solutions and quality of drawings gave evidence that students' quality of drawings did not affect the evaluation of the creativity of products. The overall inter-rater reliability among all TCPS-S tasks was .89. Then, the raters each scored tests from two different grade levels. To score Task IC, they followed the same steps that were described for the pretests.

Levels of Fidelity of Implementation (FOI)

To determine teachers' levels of FOI, which included high and low implementers, the results from another study were used (Maker & Pease, 2016). In this study, the researchers followed five steps to categorize teachers according to their levels of implementation by analyzing documents, interviews, teachers' self-ratings, supervisors' ratings, and combined ratings.

In the first step, two members of the research team reviewed the documents they collected during classroom observations: observation forms, photographs of the classroom environment, students' products, written notes, group and individual discussions with teachers, and documents that were given to the students. After analyzing these documents, each of the two research team members completed a checklist form (part of the form is presented in Appendix K). This form contains four sections based on the differentiation components of the REAPS model: content, process, product, and learning environment, with a fifth section for general comments. A 6-point scale was used for this form, with 0 = no evidence, 1 = low, 2 = below average, 3 = average, 4 = above average, 5 = high, and 6 = excellent. Later, the two members of the research team compared their ratings for each section of the checklist to establish inter-rater agreement using Cronbach's Alpha. The value of each section and across all four sections was above .75, which is considered acceptable.

In the second step, members of the research team analyzed the interview data based on the five principles of FOI (Dane, 1997; Dane & Schneider, 1998; Dusenbury et al., 2003; Moncher & Prinz,1991; O'Donnell, 2008): (a) adherence, or the extent to which a teaching unit was delivered as designed; (b) exposure, or the teacher's practices in the classroom; (c) quality of delivery, or the way the teacher delivered the REAPS model using techniques, processes, or methods prescribed; (d) participant responsiveness, or the degree to which the students participated in the program's activities; and (e) program differentiation. The two research team members listened to each teacher's interview and gave one point every time the teacher said or mentioned an idea that demonstrated one of the five principles of FOI. The interview scores were sorted from highest to lowest; a score of 20 or higher indicated that the teacher understood the REAPS model better than teachers who received fewer than 20 points.

In the third and fourth steps, two research team members compared their scores with the teachers' self-ratings (Appendix J) and the curriculum supervisors' ratings (Appendix K) to calculate the inter-rater reliability. The value of each section of the checklist and across all sections was above .75. Obtaining a high inter-rater reliability value helped the researchers to increase confidence in their findings.

In the fifth step, the research team summed the ratings for teachers across all four categories, and the scores were classified from highest to lowest. The cut-off point for distinguishing between high and low implementers was 4.5 out of 6; teachers who received an average score of 4.5 or higher were labeled high implementers, while teachers who earned less than 4.5 were considered low implementers. The scores of the high implementers ranged from 4.7 to 5.5, while the scores of the low implementers ranged from 3.6 to 4.3.

Variables

Grade levels

This served as the first independent variable. Four grade levels were included: 2, 3, 4, and 5 in Year 1 and 3, 4, 5, and 6 in Year 2. The number of participants based on grade level, classrooms, and levels of FOI in Years 1 and 2 are presented in Tables 13 and 14.

Fidelity of Implementation (FOI)

This served as the second independent variable. A categorical technique was used to divide teachers into two groups (high and low implementers) because only a small sample size of 17 classrooms in Year 1 and 16 classrooms in Year 2 was included in this analysis. The variation among teachers could be more easily determined by using this technique.

The Test of Creative Problem Solving in Science (TCPS-S)

Scores on this test were dependent variables. For Questions 1 and 2, five dependent variables were used: the total scores on TCPS-S, fluency, flexibility, elaboration, and originality. For Question 3, the three factors of TCPS-S (Maker et al., 2017) were used as dependent variables: Finding Problems, Generating Detailed Solutions, and Classifying Elements.

Table 13: *Number of Students Based on Grades and Levels of FOI Years 1 and 2*

Levels of FOI	Grades		Number of Students
	Year 1	Year 2	
Low implementation in Years 1 and 2	2	3	25
	3	4	0
	4	5	0
	5	6	7
		Total	**32**
High implementation in Years 1 and 2	2	3	8
	3	4	38
	4	5	41
	5	6	23
		Total	**110**
High implementation in Year 1 and Low implementation in Year 2	2	3	47
	3	4	0
	4	5	36
	5	6	41
		Total	**124**
Low implementation in Year 1 and High implementation in Year 2	2	3	8
	3	4	41
	4	5	0
	5	6	2
		Total	**51**

Note: FOI = Fidelity of Implementation.

Table 14: *Number of Classrooms Based on Grades and Levels of FOI Years 1 and 2*

Levels of FOI	Year 1			Year 2		
	Grades	Classrooms	Students	Grades	Classrooms	Students
Low FOI	2	2	33	3	4	72
	3	2	41	4	0	0
	4	0	0	5	2	36
	5	1	9	6	2	48
High FOI	2	3	55	3	1	16
	3	2	38	4	4	79
	4	4	77	5	2	41
	5	4	64	6	1	25

Note: FOI = Fidelity of implementation.

Data Analysis

One-Way Repeated Measures Multivariate Analysis Of Variance (MANOVA) was applied to answer the research questions. Repeated Measures MANOVA is considered an appropriate technique to test the growth in student creativity from pretest to posttest with one or more independent variables and two or more dependent variables (Frey, 2016; Tabachnick & Fidell, 2013).

Threats to Validity

Three internal and one external threat to the validity of this study were identified. Testing, instrumentation, and attrition were the major internal threats that might arise. The internal threat of testing is the possibility that the participants were familiar with the test tasks, especially when the study included pretests and posttests (Creswell, 2009). To address this threat, the duration between administering the pretest and posttest was more than one year and the test used in this study has no ceiling scores. The other internal threat to validity was instrumentation, the changes between the tasks in the pretest and posttest that might impact the scores (Creswell).

To avoid this threat, both pretests and posttests had the same tasks and questions. Attrition is withdrawal from the study for a number of reasons such as relocating away from the area in which the school was located (Miller & Hollist, 2007). To address this threat, 47 students who completed only the pretests were excluded.

For the external threat to validity, interaction of setting and treatment was the major problem. The findings of this study must be interpreted with caution because of the characteristics of the school where the study was conducted (Creswell, 2009). In the school, approximately 52% of the students were from language backgrounds other than English, with more than 50 languages represented. The participants were not randomly assigned because of the nature of the study. Thus, to address this threat, future studies are recommended in different settings for comparison with the results of this study.

CHAPTER IV: RESULTS

Three research questions guided this study to examine the influence of the levels of fidelity of implementation of the REAPS model on students' creative problem solving in science. One-Way Repeated Measures Multivariate Analysis of Variance (MANOVA) was used to answer these questions.

Research Question 1: What were the differences in students' growth in creative problem solving in science at different grade levels?

One-way Repeated Measures MANOVA was run to examine the effects of the implementation of the REAPS model on growth in students' creative problem solving at different grade levels. Five dependent variables were used (total scores on the TCPS-S, fluency, flexibility, elaboration, and originality), with one independent variable (grade levels).

Test of Assumptions

Before the data were analyzed, several assumptions were assessed: multicollinearity, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, homogeneity of variance, and normality of data. No evidence of multicollinearity was found among the dependent variables, as assessed by Pearson correlation ($r < 0.9$; Tabachnick & Fidell, 2013). A linear relationship between dependent variables was found, as assessed by scatterplot.

Univariate outliers were found in the data, as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box (Laerd Statistics, 2015). Multivariate outliers also were found in the data as assessed by Mahalanobis distance with a critical value of 20.515 when the number of dependent variables is five (Laerd Statistics; Tabachnick & Fidell, 2013). Data entry was checked, and no errors for recording scores were found. The outliers were not removed from the analyses because the critical objective of the

REAPS model was to promote creativity among students. Outliers were expected because the TCPS-S instrument was designed to identify creative students in science and it had no maximum scores.

The assumption of homogeneity of covariance matrices was violated, as assessed by Box's M test ($p < .001$). However, Pillai's Trace is considered more robust (Olsen, 1976) and is recommended especially when sample sizes are unequal, which might otherwise lead to a statistically significant Box's M result (Tabachnick & Fidell, 2013). The values of homogeneity of variance, as assessed by Levene's Test of Homogeneity of Variance, ranged from $p = .000$ to $p = .823$. In the pretest, significant homogeneity of variance was found in Total TCPS-S scores ($p = .028$), fluency ($p = .004$), elaboration ($p = .000$), and originality ($p = .000$). Small differences among the groups' variances may lead to significant homogeneity in Levene's test, especially when the sample size is large (Field, 2009). Thus, the results should be interpreted with caution (George & Mallery, 2016). Normality of data was assessed by the values of skewness and kurtosis. In skewness, the values among dependent variables ranged from -0.278 to 2.351. The acceptable value of skewness is ± 2 (George & Mallery). The values of kurtosis ranged from -0.411 to 7.515, and "[a] kurtosis greater than 10 is problematic; a kurtosis over 20 is very serious" (Acock, 2014, p. 110). Skewness and kurtosis values for pretests and posttests are presented in Table 15.

Table 15: *Skewness and Kurtosis Values for Pretests and Posttests*

Dependent Variable	Test	Grade Level	*N*	Skewness		Kurtosis	
				Statistic	*SE*	Statistic	*SE*
Total TCPS-S	Pretest	Grade 2Y1 & 3Y2	88	0.760	0.257	0.920	0.508
		Grade 3Y1 & 4Y2	79	0.717	0.271	0.882	0.535
		Grade 4Y1 & 5Y2	77	0.919	0.274	1.517	0.541
		Grade 5Y1 & 6Y2	73	0.774	0.281	1.079	0.555
	Posttest	Grade 2Y1 & 3Y2	88	0.242	0.257	-0.198	0.508
		Grade 3Y1 & 4Y2	79	0.518	0.271	0.221	0.535
		Grade 4Y1 & 5Y2	77	0.457	0.274	0.055	0.541
		Grade 5Y1 & 6Y2	73	0.830	0.281	2.096	0.555
Fluency	Pretest	Grade 2Y1 & 3Y2	88	1.116	0.257	1.532	0.508
		Grade 3Y1 & 4Y2	79	0.572	0.271	0.001	0.535
		Grade 4Y1 & 5Y2	77	1.145	0.274	2.118	0.541
		Grade 5Y1 & 6Y2	73	0.499	0.281	0.671	0.555
	Posttest	Grade 2Y1 & 3Y2	88	0.204	0.257	0.107	0.508
		Grade 3Y1 & 4Y2	79	0.192	0.271	0.370	0.535
		Grade 4Y1 & 5Y2	77	0.659	0.274	0.847	0.541
		Grade 5Y1 & 6Y2	73	0.429	0.281	0.476	0.555
Flexibility	Pretest	Grade 2Y1 & 3Y2	88	0.114	0.257	-0.316	0.508
		Grade 3Y1 & 4Y2	79	0.203	0.271	0.139	0.535
		Grade 4Y1 & 5Y2	77	0.302	0.274	-0.167	0.541
		Grade 5Y1 & 6Y2	73	0.653	0.281	0.612	0.555
	Posttest	Grade 2Y1 & 3Y2	88	-0.278	0.257	-0.028	0.508
		Grade 3Y1 & 4Y2	79	0.077	0.271	0.174	0.535
		Grade 4Y1 & 5Y2	77	0.314	0.274	-0.092	0.541
		Grade 5Y1 & 6Y2	73	0.529	0.281	1.377	0.555
Elaboration	Pretest	Grade 2Y1 & 3Y2	88	2.064	0.257	4.303	0.508
		Grade 3Y1 & 4Y2	79	1.455	0.271	1.739	0.535
		Grade 4Y1 & 5Y2	77	1.609	0.274	3.363	0.541
		Grade 5Y1 & 6Y2	73	1.489	0.281	3.384	0.555
	Posttest	Grade 2Y1 & 3Y2	88	1.284	0.257	0.956	0.508
		Grade 3Y1 & 4Y2	79	1.048	0.271	2.039	0.535
		Grade 4Y1 & 5Y2	77	-0.274	0.274	0.018	0.541
		Grade 5Y1 & 6Y2	73	0.325	0.281	-0.411	0.555

Continued

Table 15: *Skewness and Kurtosis Values for Pretests and Posttests*

Dependent Variable	Test	Grade Level	*N*	Skewness Statistic	Skewness *SE*	Kurtosis Statistic	Kurtosis *SE*
Originality	Pretest	Grade 2Y1 & 3Y2	88	0.838	0.257	1.000	0.508
		Grade 3Y1 & 4Y2	79	1.374	0.271	1.722	0.535
		Grade 4Y1 & 5Y2	77	2.351	0.274	5.110	0.541
		Grade 5Y1 & 6Y2	73	1.672	0.281	3.267	0.555
	Posttest	Grade 2Y1 & 3Y2	88	0.352	0.257	-0.208	0.508
		Grade 3Y1 & 4Y2	79	1.478	0.271	3.027	0.535
		Grade 4Y1 & 5Y2	77	1.430	0.274	2.841	0.541
		Grade 5Y1 & 6Y2	73	1.984	0.281	7.515	0.555

Note: TCPS-S = Test of Creative Problem Solving in Science; Grade 2Y1 & 3Y2 = Students who were in grade 2 in Year 1 and then entered grade 3 in Year 2; Grade 3Y1 & 4Y2 = Students who were in grade 3 in Year 1 and then entered grade 4 in Year 2; Grade 4Y1 & 5Y2 = Students who were in grade 4 in Year 1 and then entered grade 5 in Year 2; Grade 5Y1 & 6Y2 = Students who were in grade 5 in Year 1 and then entered in grade 6 in Year 2; *N* = Sample size; *SE* = Std. Error.

Results

A statistically significant interaction effect was found between growth of creative problem solving in science scores and grade levels on the combined dependent variables, $F(15, 933) = 17.475$, $p = .000$, Pillai's Trace = .658, $\eta^2 = .219$. Follow-up univariate one-way ANOVAs were performed.

A statistically significant interaction effect was found between growth in the total scores on the TCPS-S and grade levels, $F(3, 313) = 5.510$, $p = .001$, $\eta^2 = .050$. Tukey post hoc tests were run for the differences in total scores on the TCPS-S among grade levels. Students who were in grade 3 in Year 1 and grade 4 in Year 2 showed growth in their scores, which was statistically significant when compared with growth of students who were in grade 2 in Year 1 and grade 3 in Year 2 ($p = .000$). Students who were in grade 4 in Year 1 and grade 5 in Year 2 showed growth in their scores, which was statistically significant when compared with growth of students who were in grade 2 in Year 1 and grade 3 in Year 2 ($p = .027$).

A statistically significant interaction effect was found between growth in fluency scores and grade levels, $F(3, 313) = 16.899$, $p = .000$, $\eta^2 = .139$. Tukey post hoc tests were run for the differences in fluency among grade levels. Students who were in grade 3 in Year 1 and in grade 4 in Year 2 showed growth in fluency scores, which was statistically significant when compared with growth of students in two groups: students who were in grade 2 in Year 1 and grade 3 in Year 2 ($p = .000$), and students who were in grade 5 in Year 1 and grade 6 in Year 2 ($p = .033$). Students who were in grade 4 in Year 1 and grade 5 in Year 2 showed growth in fluency scores, which was statistically significant when compared with growth of students who were in grade 2 in Year 1 and grade 3 in Year 2 ($p = .000$). Students who were in grade 5 in Year 1 and grade 6 in Year 2 showed growth in fluency scores, which was statistically significant when compared with growth of students who were in grade 2 in Year 1 and grade 3 in Year 2 ($p = .000$).

A statistically significant interaction effect was found between growth in flexibility scores and grade levels, $F(3, 313) = 12.092$, $p = .000$, $\eta^2 = .104$. Tukey post hoc tests were run for the differences in flexibility between grade levels. Students who were in grade 3 in Year 1 and grade 4 in Year 2 showed growth in flexibility scores, which was statistically significant when compared with growth of students in two groups: students who were in grade 2 in Year 1 and grade 3 in Year 2 ($p = .000$), and students who were in grade 5 in Year 1 and grade 6 in Year 2 ($p = .014$). Students who were in grade 4 in Year 1 and grade 5 in Year 2 showed growth in flexibility scores, which was statistically significant when compared with growth of students who were in grade 2 in Year 1 and grade 3 in Year 2 ($p = .001$). Students who were in grade 5 in Year 1 and grade 6 in Year 2 showed growth in flexibility scores, which was statistically significant when compared with growth of students who were in grade 2 in Year 1 and grade 3 in Year 2 ($p = .007$).

A statistically significant interaction effect was found between growth in elaboration scores and grade levels, $F(3, 313) = 30.892$, $p = .000$, $\eta^2 = .228$. Tukey post hoc tests were run for the differences in elaboration between grade levels. Students who were in grade 3 in Year 1 and grade 4 in Year 2 showed growth in elaboration scores, which was statistically significant when compared with growth of students who were in grade 2 in Year 1 and grade 3 in Year 2 ($p = .000$). Students who were in grade 4 in Year 1 and grade 5 in Year 2 showed growth in elaboration scores, which was statistically significant when compared with growth of students in three groups: students who were in grade 2 in Year 1 and grade 3 in Year 2 ($p = .000$), students who were in grade 3 in Year 1 and grade 4 in Year 2 ($p = .009$), and students who were in grade 5 in Year 1 and grade 6 in Year 2 ($p = .013$). Students who were in grade 5 in Year 1 and grade 6 in Year 2 showed growth in elaboration scores, which was statistically significant when compared with growth of students who were in grade 2 in Year 1 and grade 3 in Year 2 ($p = .000$).

A statistically significant interaction effect was found between growth in originality scores and grade levels, $F(3, 313) = 6.331$, $p = .000$, $\eta^2 = .057$. Tukey post hoc tests were run for the differences in originality between grade levels. Students who were in grade 2 in Year 1 and grade 3 in Year 2 showed growth in originality scores, which was statistically significant when compared with growth of students in three groups: students who were in grade 3 in Year 1 and grade 4 in Year 2 ($p = .000$), students who were in grade 4 in Year 1 and grade 5 in Year 2 ($p = .000$), and students who were in grade 5 in Year 1 and grade 6 in Year 2 ($p = .000$). Students who were in grade 3 in Year 1 and grade 4 in Year 2 showed growth in originality scores, which was statistically significant when compared with growth of students who were in grade 4 in Year 1 and grade 5 in Year 2 ($p = .000$). Students who were in grade 5 in Year 1 and grade 6 in Year 2 showed growth in originality scores, which was statistically significant when compared with growth of students who were in grade 4 in Year 1 and grade 5 in Year 2 (p = .003). Descriptive statistics based on grade level for pretests and posttests are presented in Table 16. A summary of the results for Question 1 is presented in Figure 6.

Table 16: *Descriptive Statistics Based on Grade Level for Pretests and Posttests*

Dependent Variable	Test	Grade Level	*N*	*M*	*SD*	Min.	Max.
Total TCPS-S	Pretest	Grade 2Y1 & 3Y2	88	62.97	27.65	5.00	143.00
		Grade 3Y1 & 4Y2	79	86.80	35.95	14.00	194.00
		Grade 4Y1 & 5Y2	77	71.33	25.59	22.00	164.67
		Grade 5Y1 & 6Y2	73	77.05	25.27	29.00	158.00
	Posttest	Grade 2Y1 & 3Y2	88	84.78	28.19	25.00	156.50
		Grade 3Y1 & 4Y2	79	95.30	29.95	38.50	188.50
		Grade 4Y1 & 5Y2	77	97.92	32.93	29.50	201.50
		Grade 5Y1 & 6Y2	73	87.64	29.36	18.00	199.00
Fluency	Pretest	Grade 2Y1 & 3Y2	88	17.33	9.57	1.00	49.00
		Grade 3Y1 & 4Y2	79	30.59	13.43	3.00	67.00
		Grade 4Y1 & 5Y2	77	26.57	10.33	8.00	63.00
		Grade 5Y1 & 6Y2	73	28.23	9.95	6.00	61.00
	Posttest	Grade 2Y1 & 3Y2	88	19.78	7.03	4.00	39.00
		Grade 3Y1 & 4Y2	79	24.09	6.91	8.00	46.00
		Grade 4Y1 & 5Y2	77	22.19	7.24	10.00	47.00
		Grade 5Y1 & 6Y2	73	19.81	6.52	4.00	36.00
Flexibility	Pretest	Grade 2Y1 & 3Y2	88	25.47	9.56	3.00	51.00
		Grade 3Y1 & 4Y2	79	34.33	11.90	9.00	63.00
		Grade 4Y1 & 5Y2	77	33.35	9.81	12.00	57.00
		Grade 5Y1 & 6Y2	73	35.01	10.25	15.00	66.00
	Posttest	Grade 2Y1 & 3Y2	88	33.65	9.84	9.00	57.00
		Grade 3Y1 & 4Y2	79	42.00	11.38	12.00	72.00
		Grade 4Y1 & 5Y2	77	36.04	10.95	15.00	66.00
		Grade 5Y1 & 6Y2	73	32.92	10.23	9.00	69.00
Elaboration	Pretest	Grade 2Y1 & 3Y2	88	4.30	6.85	0.00	33.00
		Grade 3Y1 & 4Y2	79	10.44	11.00	0.00	45.00
		Grade 4Y1 & 5Y2	77	7.01	8.65	0.00	42.00
		Grade 5Y1 & 6Y2	73	3.86	2.94	0.00	15.00

Continued

Table 16: *Descriptive Statistics Based on Grade Level for Pretests and Posttests*

Dependent Variable	Test	Grade Level	*N*	*M*	*SD*	Min.	Max.
	Posttest	Grade 2Y1 & 3Y2	88	9.89	10.65	0.00	42.00
		Grade 3Y1 & 4Y2	79	16.52	12.20	0.00	66.00
		Grade 4Y1 & 5Y2	77	27.78	13.90	0.00	66.00
		Grade 5Y1 & 6Y2	73	23.30	10.69	0.00	42.00
Originality	Pretest	Grade 2Y1 & 3Y2	88	13.73	8.50	0.00	40.00
		Grade 3Y1 & 4Y2	79	8.29	7.86	0.00	32.00
		Grade 4Y1 & 5Y2	77	0.90	1.70	0.00	8.00
		Grade 5Y1 & 6Y2	73	6.78	6.88	0.00	34.00
	Posttest	Grade 2Y1 & 3Y2	88	17.92	8.03	0.00	37.00
		Grade 3Y1 & 4Y2	79	9.49	8.01	0.00	40.00
		Grade 4Y1 & 5Y2	77	8.42	7.00	0.00	35.00
		Grade 5Y1 & 6Y2	73	8.67	7.64	0.00	47.00

Note: TCPS-S = Test of Creative Problem Solving in Science; Grade 2Y1 & 3Y2 = Students who were in grade 2 in Year 1 and then entered grade 3 in Year 2; Grade 3Y1 & 4Y2 = Students who were in grade 3 in Year 1 and then entered grade 4 in Year 2; Grade 4Y1 & 5Y2 = Students who were in grade 4 in Year 1 and then entered grade 5 in Year 2; Grade 5Y1 & 6Y2 = Students who were in grade 5 in Year 1 and then entered grade 6 in Year 2; *N* = Sample size; *SD* = Standard Deviation; Min. = Minimum; Max. = Maximum.

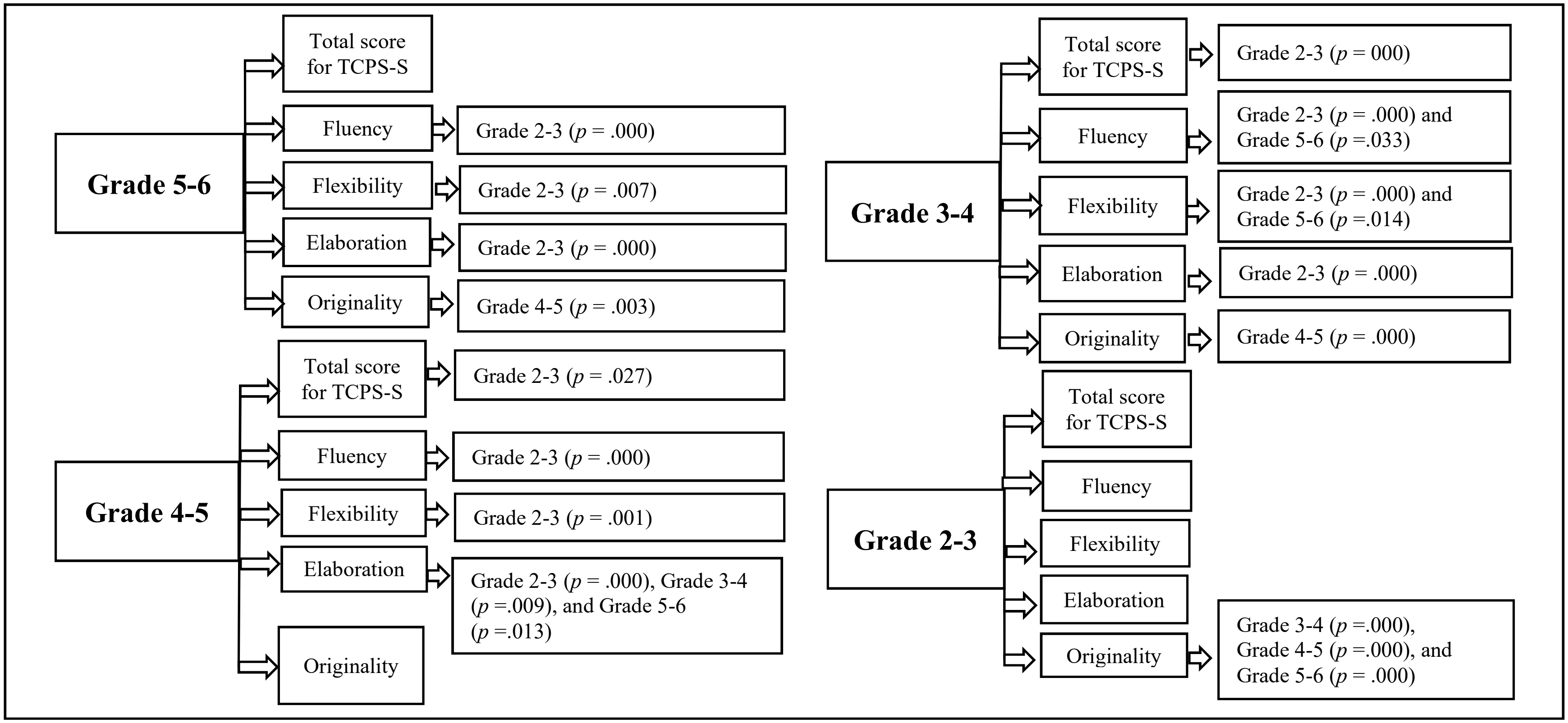

Figure 6: A Summary of the Results for Question 1

Note: Grade 5-6 = Students who were in grade 5 in Year 1 and then entered grade 6 in Year 2; Grade 4-5 = Students who were in grade 4 in Year 1 and then entered grade 5 in Year 2; Grade 3-4 = Students who were in grade 3 in Year 1 and then entered grade 4 in Year 2; Grade 2-3 = Students who were in grade 2 in Year 1 and then entered grade 3 in Year 2

Only statistically significant differences were included in this figure. It illustrates the growth of creative problem solving in science among grade levels. The top left figure, for example, showed a comparison of students who were in grade 5-6 with three other groups. This group showed statistically significant differences from students who were in grade 2-3 in fluency, flexibility and elaboration. Also, they showed statistically significant differences from students who were in grade 4-5 in originality.

Research Question 2: What were the differences in growth of creativity of students who were in classrooms of teachers who were (a) at a high level of fidelity for two years (HF), (b) at a low level of fidelity for two years (LF), and (c) at mixed levels of fidelity (i.e., high level of fidelity in Year 1 but low level of fidelity in Year 2 [HLF], and low level of fidelity in Year 1 but high level of fidelity in Year 2 [LHF])?

One-Way Repeated Measures MANOVA was used to investigate the impact of the implementation of the REAPS model on growth in students' creative problem solving at different levels of fidelity of implementation (FOI). Five dependent variables were used (total scores on the TCPS-S, fluency, flexibility, elaboration, and originality), with one independent variable (levels of FOI).

Test of Assumptions

Several assumptions were assessed before running the analysis for Question 2: multicollinearity, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, homogeneity of variance, and normality of data. No evidence of multicollinearity was found among dependent variables, as assessed by Pearson correlation ($r < 0.9$; Tabachnick & Fidell, 2013). A linear relationship between dependent variables was found, as assessed by scatterplot.

Univariate outliers, as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box, were found in the data (Laerd Statistics, 2015). Multivariate outliers also were found, as assessed by Mahalanobis distance, with a critical value of 20.515 for five dependent variables (Laerd Statistics; Tabachnick & Fidell, 2013). No errors for recording scores were found in data entry. The outliers were kept in the analyses because the purpose of the intervention was to increase creativity among participants. Outliers were anticipated because the TCPS-S instrument was developed to identify creative students in science and it had no maximum scores.

The assumption of homogeneity of covariance matrices was not met, as assessed by Box's M test ($p < .001$). However, Pillai's Trace was used because it is considered more robust (Olsen, 1976) and is suggested when sample sizes are unequal, which might otherwise lead to a statistically significant Box's M result (Tabachnick & Fidell, 2013). The values of homogeneity of variance, as assessed by Levene's Test of Homogeneity of Variance, ranged from $p = .000$ to $p = .580$. A significant homogeneity of variance was found in Total TCPS-S posttest scores ($p = .006$), in elaboration on the pretest ($p = .042$) and posttest ($p = .000$), and in originality on the pretest ($p = .020$) and posttest ($p = .000$). Thus, the results should be interpreted carefully (George & Mallery, 2016). Normality of data was assessed by examining the values of skewness and kurtosis. The skewness values among dependent variables ranged from -0.24 to 2.262. The acceptable value of skewness is ± 2 (George & Mallery). The values of kurtosis ranged from -0.89 to 7.329; "[a] kurtosis greater than 10 is problematic; a kurtosis over 20 is very serious" (Acock, 2014, p. 110). Skewness and kurtosis values for pretests and posttests are presented in Table 17.

Table 17: *Skewness and Kurtosis Values for Pretests and Posttests*

Dependent Variable	Test	Levels of Fidelity of Implementation	*N*	Skewness		Kurtosis	
				Statistic	*SE*	Statistic	*SE*
Total TCPS-S	Pretest	Low Y1 & 2	32	1.034	0.414	2.027	0.809
		High Y1 & 2	110	1.011	0.23	1.807	0.457
		High Y1 & Low Y2	124	0.563	0.217	0.578	0.431
		Low Y1 & High Y2	51	1.011	0.333	1.944	0.656
	Posttest	Low Y1 & 2	32	0.24	0.414	-0.411	0.809
		High Y1 & 2	110	0.498	0.23	0.106	0.457
		High Y1 & Low Y2	124	0.442	0.217	0.386	0.431
		Low Y1 & High Y2	51	0.331	0.333	-0.176	0.656

Continued

Table 17: *Skewness and Kurtosis Values for Pretests and Posttests*

Dependent Variable	Test	Levels of Fidelity of Implementation	*N*	Skewness		Kurtosis	
				Statistic	*SE*	Statistic	*SE*
Fluency	Pretest	Low Y1 & 2	32	0.667	0.414	0.124	0.809
		High Y1 & 2	110	1.033	0.23	1.301	0.457
		High Y1 & Low Y2	124	0.569	0.217	0.372	0.431
		Low Y1 & High Y2	51	0.314	0.333	-0.356	0.656
	Posttest	Low Y1 & 2	32	0.515	0.414	0.459	0.809
		High Y1 & 2	110	0.248	0.23	-0.07	0.457
		High Y1 & Low Y2	124	0.389	0.217	0.515	0.431
		Low Y1 & High Y2	51	0.218	0.333	0.367	0.656
Flexibility	Pretest	Low Y1 & 2	32	0.294	0.414	-0.549	0.809
		High Y1 & 2	110	0.154	0.23	0.138	0.457
		High Y1 & Low Y2	124	0.286	0.217	0.346	0.431
		Low Y1 & High Y2	51	0.419	0.333	0.144	0.656
	Posttest	Low Y1 & 2	32	-0.24	0.414	-0.352	0.809
		High Y1 & 2	110	0.106	0.23	-0.27	0.457
		High Y1 & Low Y2	124	0.167	0.217	0.329	0.431
		Low Y1 & High Y2	51	0.296	0.333	-0.329	0.656
Elaboration	Pretest	Low Y1 & 2	32	1.863	0.414	3.722	0.809
		High Y1 & 2	110	1.975	0.23	4.095	0.457
		High Y1 & Low Y2	124	2.262	0.217	7.329	0.431
		Low Y1 & High Y2	51	1.841	0.333	3.813	0.656
	Posttest	Low Y1 & 2	32	1.387	0.414	1.431	0.809
		High Y1 & 2	110	0.053	0.23	-0.701	0.457
		High Y1 & Low Y2	124	0.285	0.217	-0.88	0.431
		Low Y1 & High Y2	51	1.054	0.333	2.078	0.656
Originality	Pretest	Low Y1 & 2	32	1.337	0.414	2.772	0.809
		High Y1 & 2	110	1.503	0.23	2.307	0.457
		High Y1 & Low Y2	124	1.184	0.217	0.948	0.431
		Low Y1 & High Y2	51	1.387	0.333	1.566	0.656
	Posttest	Low Y1 & 2	32	0.303	0.414	-0.89	0.809
		High Y1 & 2	110	1.098	0.23	1.369	0.457
		High Y1 & Low Y2	124	0.911	0.217	1.223	0.431
		Low Y1 & High Y2	51	0.968	0.333	1.137	0.656

Note: TCPS-S = Test of Creative Problem Solving in Science; Low Y1 & 2 = Students who were with Low implementers in Years 1 and 2; High Y1 & 2 = Students who were with High implementers in Years 1 and 2; High Y1 & Low Y2 = Students who were with High implementers in Year 1 and then with Low implementers in Year 2; Low Y1 & High Y2 = Students who were with Low implementers in Year 1 and then with High implementers in Year 2; *N* = Sample size; *SE* = Std. Error.

Results

A statistically significant interaction effect was found between growth of creative problem solving in science scores and levels of FOI on the combined dependent variables, $F(15, 933) = 3.353$, $p = .000$, Pillai's Trace = .153, $\eta^2 = .051$. Follow-up univariate one-way ANOVAs were performed.

A statistically significant interaction was found between growth in flexibility scores and levels of FOI, $F(3, 313) = 5.621$, $p = .001$, $\eta^2 = .051$. Tukey post-hoc tests were run for the differences among levels of FOI. Students who were with low implementers in Year 1 and then were with high implementers in Year 2 showed growth in flexibility scores, which was statistically significant when compared with growth of students in three groups: students who were with low implementers in Years 1 and 2 ($p = .003$), students who were with high implementers in Years 1 and 2 ($p = .026$), and students who were with high implementers in Year 1 and then with low implementers in Year 2 ($p = .019$).

A statistically significant interaction effect was found between growth in elaboration scores and levels of FOI, $F(3, 313) = 5.597$, $p = .001$, $\eta^2 = .051$. Tukey post-hoc tests were run for the differences among levels of FOI. Students who were with high implementers in Years 1 and 2 showed growth in elaboration scores, which was statistically significant when compared with growth of students who were with low implementers in Years 1 and 2 ($p = .038$).

No statistically significant interaction effect was found between growth in creative problem solving in science scores and levels of FOI in total scores for TCPS-S, $F(3, 313) = 1.866$, $p = .135$, $\eta^2 =.018$; fluency, $F(3, 313) = 2.386$, $p = .069$, $\eta^2 =.022$; and originality, $F(3, 313) = 1.612$, $p = .186$, $\eta^2 =.015$. Descriptive statistics for each dependent variable are presented in Table 18. A summary of the results for Question 2 is presented in Figure 7.

Table 18: *Descriptive Statistics Based on Level of Implementation for Pretests and Posttests*

Dependent Variable	Test	Levels of Fidelity of Implementation	*N*	*M*	*SD*	Min.	Max.
Total TCPS-S	Pretest	Low Y1 & 2	32	65.94	26.43	19.00	140.67
		High Y1 & 2	110	74.65	30.24	14.00	194.00
		High Y1 & Low Y2	124	73.69	29.33	5.00	164.67
		Low Y1 & High Y2	51	79.52	33.79	18.00	190.67
	Posttest	Low Y1 & 2	32	84.05	29.13	25.00	147.00
		High Y1 & 2	110	85.76	23.51	38.50	149.50
		High Y1 & Low Y2	124	93.89	34.16	18.00	201.50
		Low Y1 & High Y2	51	101.23	32.26	44.00	188.50
Fluency	Pretest	Low Y1 & 2	32	18.94	9.11	3.00	41.00
		High Y1 & 2	110	26.61	12.28	3.00	67.00
		High Y1 & Low Y2	124	25.04	11.94	1.00	63.00
		Low Y1 & High Y2	51	27.67	12.24	6.00	57.00
	Posttest	Low Y1 & 2	32	19.03	6.72	7.00	36.00
		High Y1 & 2	110	21.04	6.11	8.00	39.00
		High Y1 & Low Y2	124	20.98	7.59	4.00	47.00
		Low Y1 & High Y2	51	25.00	7.30	9.00	46.00
Flexibility	Pretest	Low Y1 & 2	32	27.38	9.39	9.00	45.00
		High Y1 & 2	110	32.45	10.35	9.00	63.00
		High Y1 & Low Y2	124	31.96	11.55	3.00	66.00
		Low Y1 & High Y2	51	32.71	12.09	9.00	63.00
	Posttest	Low Y1 & 2	32	34.22	9.87	12.00	51.00
		High Y1 & 2	110	34.75	9.63	12.00	57.00
		High Y1 & Low Y2	124	35.06	11.45	9.00	69.00
		Low Y1 & High Y2	51	43.00	11.90	21.00	72.00
Elaboration	Pretest	Low Y1 & 2	32	6.09	7.79	0.00	33.00
		High Y1 & 2	110	7.25	9.26	0.00	42.00
		High Y1 & Low Y2	124	4.98	6.72	0.00	42.00
		Low Y1 & High Y2	51	8.12	9.72	0.00	45.00
	Posttest	Low Y1 & 2	32	11.44	11.57	0.00	42.00
		High Y1 & 2	110	19.42	11.50	0.00	42.00
		High Y1 & Low Y2	124	21.02	15.57	0.00	66.00
		Low Y1 & High Y2	51	17.76	13.16	0.00	66.00

Continued

Table 18: *Descriptive Statistics Based on Level of Implementation for Pretests and Posttests*

Dependent Variable	Test	Levels of Fidelity of Implementation	*N*	*M*	*SD*	Min.	Max.
Originality	Pretest	Low Y1 & 2	32	10.84	8.65	0.00	40.00
		High Y1 & 2	110	5.53	6.67	0.00	32.00
		High Y1 & Low Y2	124	8.63	9.12	0.00	39.00
		Low Y1 & High Y2	51	7.88	8.05	0.00	32.00
	Posttest	Low Y1 & 2	32	15.88	10.02	0.00	36.00
		High Y1 & 2	110	7.56	6.12	0.00	30.00
		High Y1 & Low Y2	124	13.35	8.90	0.00	47.00
		Low Y1 & High Y2	51	12.02	9.05	0.00	40.00

Note: TCPS-S = Test of Creative Problem Solving in Science; Low Y1 & 2 = Students who were with Low implementers in Years 1 and 2; High Y1 & 2 = Students who were with High implementers in Years 1 and 2; High Y1 & Low Y2 = Students who were with High implementers in Year 1 and then with Low implementers in Year 2; Low Y1 & High Y2 = Students who were with Low implementers in Year 1 and then with High implementers in Year 2; *N* = Sample size; *SD* = Standard Deviation; Min. = Minimum; Max. = Maximum.

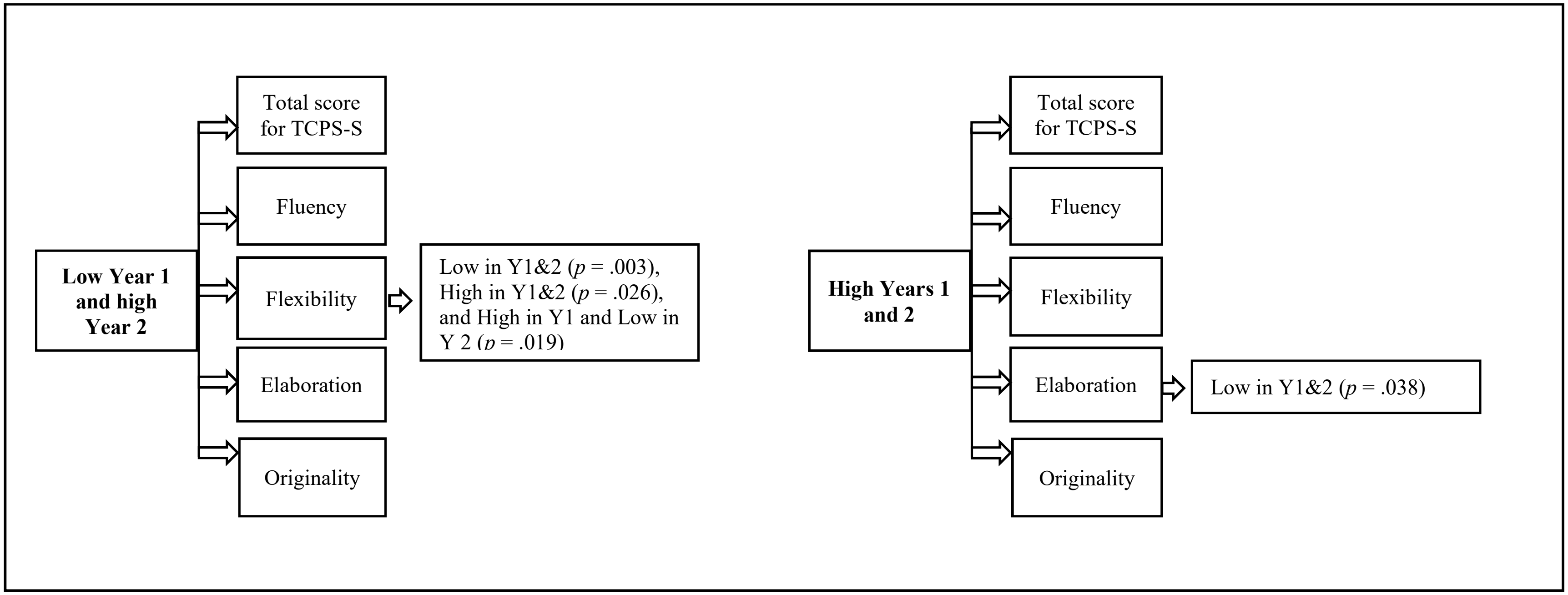

Figure 7: A Summary of the Results for Question 2

Note: Low Y1 & 2 = Students who were with Low implementers in Years 1 and 2; High Y1 & 2 = Students who were with High implementers in Years 1 and 2; High Y1 & Low Y2 = Students who were with High implementers in Year 1 and then with Low implementers in Year 2; Low Y1 & High Y2 = Students who were with Low implementers in Year 1 and then with High implementers in Year 2

Only statistically significant differences were included in this figure. It illustrates the growth of creative problem solving in science among levels of FOI. The right figure, for example, showed a comparison of students who were with high implementers in Years 1 and 2 with the other three groups. This group showed statistically significant differences from only one group (students who were with low implementers in Years 1 and 2) in elaboration. However, no differences were found between this group and the other three groups in the total scores of the TCPS-S, fluency, flexibility, and originality.

Research Question 3: Which factors of the Test of Creative Problem Solving in Science (TCPS-S) were most affected by the level of fidelity of implementation of the teachers?

One-Way Repeated Measures MANOVA was used to examine which factors of the TCPS-S were most affected by the level of fidelity of implementation of the teachers. Three dependent variables were used (Finding Problems, Generating Detailed Solutions, and Classifying Elements), with one independent variable (levels of FOI).

Test of Assumptions

Several assumptions were examined before analyzing data for Question 3: multicollinearity, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, homogeneity of variance, and normality of data. No evidence of multicollinearity was found among dependent variables, as assessed by Pearson correlation ($r < 0.9$; Tabachnick & Fidell, 2013). A linear relationship between dependent variables was found, as assessed by scatterplot.

Univariate outliers, as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box, were found in the data (Laerd Statistics, 2015). Multivariate outliers also were found, as assessed by Mahalanobis distance, with a critical value of 16.27 for three dependent variables (Laerd Statistics; Tabachnick & Fidell, 2013). No errors for recording scores were found in the data entry. The outliers were kept in the analyses because the purpose of the intervention was to increase creativity among participants. Outliers were anticipated because the TCPS-S instrument was developed to identify creative students in science and it had no maximum scores.

The assumption of homogeneity of covariance matrices was violated, as assessed by Box's M test ($p < .001$). However, Pillai's Trace was used because it is considered more robust

(Olsen, 1976) and is recommended when sample sizes are unequal, which might otherwise lead to a statistically significant Box's M result (Tabachnick & Fidell). The values of homogeneity of variance, as assessed by Levene's Test of Homogeneity of Variance, ranged from p = .000 to p = .487. A significant homogeneity of variance was found in Finding Problems on the pretest (p = .030), in Generating Detailed Solutions on the pretest (p = .014) and posttest (p = .011), and in Classifying Elements on the pretest (p = .044) and posttest (p = .000). Thus, the results should be interpreted with caution (George & Mallery, 2016). Normality of data was assessed by examining the values of skewness and kurtosis. The skewness values among dependent variables ranged from -0.063 to 1.726. The adequate value of skewness is ± 2 (George & Mallery). The values of kurtosis ranged from -0.699 to 5.135 and "[a] kurtosis greater than 10 is problematic; a kurtosis over 20 is very serious" (Acock, 2014, p. 110). Skewness and kurtosis values for pretests and posttests are presented in Table 19.

Table 19: *Skewness and Kurtosis Values Based on Levels of Fidelity of Implementation for Pretests and Posttests*

Dependent Variable	Test	Levels of Fidelity of Implementation	*N*	Skewness		Kurtosis	
				Statistic	*SE*	Statistic	*SE*
Finding Problems	Pretest	Low Y1 & 2	32	0.079	0.414	-0.455	0.809
		High Y1 & 2	110	0.427	0.23	0.106	0.457
		High Y1 & Low Y2	124	0.892	0.217	1.816	0.431
		Low Y1 & High Y2	51	0.652	0.333	1.266	0.656
	Posttest	Low Y1 & 2	32	0.575	0.414	0.481	0.809
		High Y1 & 2	110	0.571	0.23	0.124	0.457
		High Y1 & Low Y2	124	0.63	0.217	0.099	0.431
		Low Y1 & High Y2	51	0.792	0.333	1.312	0.656
Generating Detailed Solutions	Pretest	Low Y1 & 2	32	0.409	0.414	0.385	0.809
		High Y1 & 2	110	0.452	0.23	-0.577	0.457
		High Y1 & Low Y2	124	0.826	0.217	1.063	0.431
		Low Y1 & High Y2	51	-0.063	0.333	0.032	0.656

Continued

Table 19: *Skewness and Kurtosis Values Based on Levels of Fidelity of Implementation for Pretests and Posttests*

Dependent Variable	Test	Levels of Fidelity of Implementation	*N*	Skewness		Kurtosis	
				Statistic	*SE*	Statistic	*SE*
	Posttest	Low Y1 & 2	32	0.77	0.414	0.892	0.809
		High Y1 & 2	110	0.611	0.23	1.218	0.457
		High Y1 & Low Y2	124	0.736	0.217	0.951	0.431
		Low Y1 & High Y2	51	0.351	0.333	-0.403	0.656
Classifying Elements	Pretest	Low Y1 & 2	32	1.594	0.414	2.944	0.809
		High Y1 & 2	110	1.367	0.23	2.296	0.457
		High Y1 & Low Y2	124	0.774	0.217	0.632	0.431
		Low Y1 & High Y2	51	1.726	0.333	5.135	0.656
	Posttest	Low Y1 & 2	32	0.97	0.414	0.194	0.809
		High Y1 & 2	110	0.225	0.23	-0.415	0.457
		High Y1 & Low Y2	124	0.412	0.217	-0.699	0.431
		Low Y1 & High Y2	51	0.779	0.333	0.994	0.656

Note: Low Y1 & 2 = Students who were with Low implementers in Years 1 and 2; High Y1 & 2 = Students who were with High implementers in Years 1 and 2; High Y1 & Low Y2 = Students who were with High implementers in Year 1 and then with Low implementers in Year 2; Low Y1 & High Y2 = Students who were with Low implementers in Year 1 and then with High implementers in Year 2; *N* = Sample size; *SE* = Std. Error.

Results

A statistically significant interaction effect was found between levels of FOI on the combined dependent variables, $F(9, 939) = 3.682$, $p = .000$, Pillai's Trace = .102, $\eta^2 = .034$. Follow-up univariate one-way ANOVAs were performed.

A statistically significant interaction was found between growth of scores in the Generating Detailed Solutions factor and levels of FOI, $F(3, 313) = 5.762$, $p = .001$, $\eta^2 = .052$. Tukey post hoc tests were conducted for the differences among levels of FOI. Although the Tukey post hoc tests did not show any significant p value among the levels of FOI, students who were with low implementers in Year 1 and high implementers in Year 2 showed higher growth scores in the Generating Detailed Solutions factor than students in the other three groups: low implementers in Years 1 and 2, high implementers in Years 1 and 2, and high implementers in Year 1 and Low implementers in Year 2.

A statistically significant interaction was found between growth of scores in the Classifying Elements factor and levels of FOI, $F(3, 313) = 3.455$, $p = .017$, $\eta^2 = .032$. Tukey post hoc tests were conducted for the differences among levels of FOI. Students who were with high implementers in Years 1 and 2 showed growth in Classifying Elements scores, which was statistically significant when compared with growth of students who were with low implementers in Years 1 and 2 ($p = .012$). Students who were with low implementers in Year 1 and high implementers in Year 2 showed growth in Classifying Elements scores, which was statistically significant when compared with growth of students in two groups: students who were with low implementers in Years 1 and 2 ($p = .000$), and students who were with high implementers in Year 1 and low implementers in Year 2 ($p = .039$).

No statistically significant differences were found among the four groups in the Finding Problems factor, $F(3, 313) = 2.489$, $p = .060$, $\eta^2 =.023$. Descriptive statistics for each dependent variable are presented in Table 20. A summary of the results for Question 3 is presented in Figure 8.

Table 20: *Descriptive Statistics Based on Level of Implementation for Pretests and Posttests*

Dependent Variable	Test	Levels of Fidelity of Implementation	*N*	*M*	*SD*	Min.	Max.
Finding Problems	Pretest	Low Y1 & 2	32	27.88	11.60	7.00	54.00
		High Y1 & 2	110	22.36	9.10	0.00	47.00
		High Y1 & Low Y2	124	26.79	12.73	4.00	80.00
		Low Y1 & High Y2	51	23.41	11.19	4.00	59.00
	Posttest	Low Y1 & 2	32	36.56	16.43	9.00	82.00
		High Y1 & 2	110	31.43	12.97	9.00	73.00
		High Y1 & Low Y2	124	34.29	14.32	7.00	77.00
		Low Y1 & High Y2	51	37.82	14.17	12.00	79.00

Continued

Table 20: *Descriptive Statistics Based on Level of Implementation for Pretests and Posttests*

Dependent Variable	Test	Levels of Fidelity of Implementation	*N*	*M*	*SD*	Min.	Max.
Generating Detailed Solutions	Pretest	Low Y1 & 2	32	12.47	6.10	1.00	29.00
		High Y1 & 2	110	15.53	8.07	1.00	33.67
		High Y1 & Low Y2	124	16.48	8.46	1.00	45.00
		Low Y1 & High Y2	51	13.37	5.96	1.00	26.67
	Posttest	Low Y1 & 2	32	17.73	7.98	5.00	40.50
		High Y1 & 2	110	15.69	6.86	1.00	39.50
		High Y1 & Low Y2	124	18.19	8.09	1.00	48.00
		Low Y1 & High Y2	51	19.95	9.82	4.00	43.50
Classifying Elements	Pretest	Low Y1 & 2	32	25.59	17.02	0.00	76.00
		High Y1 & 2	110	36.75	23.27	0.00	127.00
		High Y1 & Low Y2	124	30.42	17.20	0.00	85.00
		Low Y1 & High Y2	51	42.75	26.85	8.00	141.00
	Posttest	Low Y1 & 2	32	29.75	18.38	0.00	71.00
		High Y1 & 2	110	38.65	15.22	0.00	72.00
		High Y1 & Low Y2	124	41.40	22.01	0.00	100.00
		Low Y1 & High Y2	51	43.45	22.02	8.00	115.00

Note: TCPS-S = Test of Creative Problem Solving in Science; Low Y1 & 2 = Students who were with Low implementers in Years 1 and 2; High Y1 & 2 = Students who were with High implementers in Years 1 and 2; High Y1 & Low Y2 = Students who were with High implementers in Year 1 and then with Low implementers in Year 2; Low Y1 & High Y2 = Students who were with Low implementers in Year 1 and then with High implementers in Year 2; *N* = Sample size; *SD* = Standard Deviation; Min. = Minimum; Max. = Maximum.

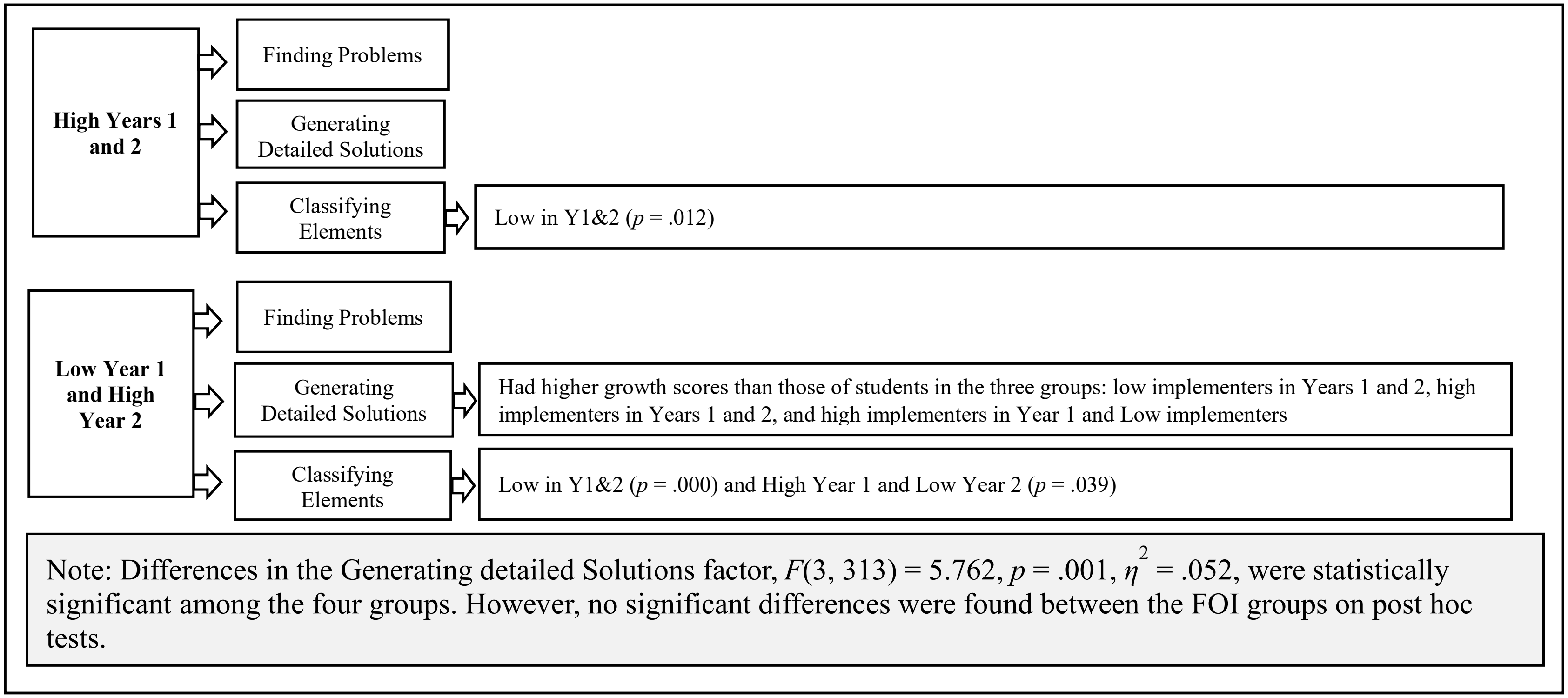

Figure 8: A Summary of the Results for Question 3

Note: Low Y1 & 2 = Students who were with Low implementers in Years 1 and 2; High Y1 & 2 = Students who were with High implementers in Years 1 and 2; High Y1 & Low Y2 = Students who were with High implementers in Year 1 and then with Low implementers in Year 2; Low Y1 & High Y2 = Students who were with Low implementers in Year 1 and then with High implementers in Year 2

Only statistically significant differences were included in this figure. It illustrates the growth of the factors of the Test of Creative Problem Solving in Science (TCPS-S) among levels of FOI. The top figure, for example, showed a comparison of students who were with high implementers in Years 1 and 2 with the other three groups. This group showed statistically significant differences from only one group (students who were with low implementers in Years 1 and 2) in the Classifying Elements factor. However, no differences were found between this group and the other three groups in the Finding Problems and Generating Detailed Solutions factors.

CHAPTER V: DISCUSSION

The purpose of this study was to determine the effect of the level of fidelity of implementation (FOI) of the Real Engagement in Active Problem Solving (REAPS) model on creative problem solving in science. Specifically, the purpose was to examine the relationships between changes in students' growth in creative problem solving in science and the level of FOI of the REAPS model.

Discussion of Findings

Discussion of Results for Research Question 1

The purpose of this question was to investigate the influence of the REAPS model on students' creative problem solving in science at different grade levels. The REAPS model, in general, had a positive effect on students' creativity in science at all grade levels. Students' scores on the Test of Creative Problem Solving in Science (TCPS-S) increased from pretests to posttests in total scores on the TCPS-S, flexibility, elaboration, and originality. These results are similar to the findings of a study by Alhusaini (2016) in which his research was implemented in the same school as the current study. Alhusaini investigated the differences in students' general creativity and creativity in science after exposure to REAPS in two different durations, long and short. Students who were exposed to the REAPS model for a long duration showed greater increases in their scientific creativity than students who were exposed to the intervention for a short duration. However, no statistically significant differences were found between the two groups in general creativity. In the current study, the students showed growth in their creativity in science after participating in the REAPS model for two years, which may be considered as a long duration. The other important aspect of the two studies is that developing domain-specific creativity may take longer than developing general creativity. In addition, the school environment may be one in which general creativity is developed, but high FOI and use of REAPS is needed for optimum development of domain-specific creativity.

The growth of students' scores supports the theories and ideas related to domain-specific expertise (Amabile, 1983; 1996; 2013; Sternberg, 1999). In these theories, some factors play significant roles in helping people to be creative individuals, such as knowledge the individuals gain about the particular domain, the individual expertise developed over time, and skills that are essential for the domain. Students developed five skills that related to domain-specific expertise (Sternberg, 1999): metacognitive skills, learning skills, thinking skills, knowledge, and motivation. For example, students learned systematic approaches for acquiring knowledge, such as distinguishing between relevant and irrelevant information that could be used to study phenomena and incorporating new knowledge into previous knowledge. The students used critical thinking skills, creative thinking skills, and practical thinking skills to solve real-world problems. For instance, they used TASC wheel steps such as Gather/Organize, Identify, Generate, Decide, Implement, and Evaluate to solve scientific problems in creative ways.

The findings of Question 1 challenge some misunderstandings of creativity: very few individuals are creative; creative ideas come to people quickly and without any effort; creative ideas should not be built from other previous ideas; and creative ideas come from experts, so novice individuals cannot be creative (Burkus, 2014; Isaksen et al., 2011; NACCCE, 1999). During the implementation of REAPS, students were exposed to a variety of scientific problems that were part of their local environment. The teachers helped students to develop the skills and strategies needed to solve these problems in creative ways. The growth of creative problem solving in science from pretest to posttest provides evidence that creativity can be taught and students need time to develop creative skills.

Students who were in grade 3 in Year 1 and grade 4 in Year 2 showed greater increases in their creative problem solving in science than students at other grade levels. These students showed greater growth in flexibility compared with all grade levels, and they also showed greater growth in fluency than students at two different grade levels (i.e., students who were in grade 2 in Year 1 and grade 3 in Year 2, and also students who were in grade 5 in Year 1 and grade 6 in Year 2). These results are similar to and different from other studies (Jo & Maker, 2011; Maker et al., 2008; Sak & Maker, 2006; Torrance, 1968).

The similarity is that levels of FOI play an important role in the growth of students' creativity (Jo & Maker, 2011; Maker et al., 2008). One possible explanation is that this group of students, who were in grade 3 in Year 1 and grade 4 in Year 2, showed greater increases in their creativity in science because they were with high implementers for two years or with low implementers in Year 1 and then with high implementers in Year 2. None of them were with low implementers for two consecutive years or with high implementers in Year 1 and then with low implementers in Year 2. The influence of FOI on the growth of students' creativity in science is discussed in detail in the section on Question 2.

In contrast, these results are different from the findings of other studies in which grade levels affected students' growth in creativity (Jo & Maker, 2011; Maker et al., 2008; Sak & Maker, 2006; Torrance, 1968). For example, Maker et al., (2008) investigated the influence of the use of the Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses (DISCOVER) curriculum model on students' general creativity as measured by the Test of Creative Thinking-Drawing Production (TCT-DP). The participants were from kindergarten to grade 6 in four schools. The researchers found that students' creativity increased from year to year with no significant peaks or slumps with the exception of grade 6 in which students showed a significant increase in their general creativity. However, the students were from ethnic groups (Native American and Hispanic) that demonstrate strengths in test domain (spatial) which may affected the findings. Another example is the study by Sak and Maker (2006) in which they investigated the relationship between age, grade levels, and domain-specific expertise on students' creative problem solving in mathematics. They used fluency, flexibility, elaboration, and originality to measure the growth of creativity. When the researchers examined the relationship between creativity and grade levels, they found that students did not show slumps or peaks in their fluency, flexibility, elaboration, and originality abilities from grade 1 to 5 with the exception of students in grade 4 in which they had a slight decrease in fluency. However, the levels of FOI were not observed in their study, which could explain these differences. Jo and Maker (2011) examined the impact of the implementation of the DISCOVER curriculum model on students' creativity in mathematics. They used data from same schools that Sak and Maker did. However, the participants of Jo's and Maker's study were from grade 1 to 5 and the researchers used the levels of fidelity of implementation. The researchers found that students' creativity in mathematics increased from year to year with no peaks or slumps. Finally, Torrance (1968) found in his longitudinal study that creativity among students in grade 4 decreased compared with students in grades 3 and 5.

One of the most notable findings was that students in the lower grades (i.e., students who were in grade 2 in Year 1 and grade 3 in Year 2) showed greater growth in originality than students at other grade levels. These results may give evidence that the REAPS model is beneficial when implemented with young children for any educational system whose policymakers aim to promote scientific creativity among students. These results also give evidence that the ability to create original responses is not associated with acquisition of knowledge. For example, students at the upper grade levels (i.e., grades 4 to 6 in Year 2) showed greater growth in elaboration scores than lower grade students. This component of creativity probably depends more on linguistic and cognitive abilities than does originality. These findings may support the idea that formal education plays a role in limiting original ideas among older students (Urban, 1991) because students tend to encounter new stresses in school (Torrance, 1977) such as "classroom etiquette and peer pressure" (Kim, 2011, p. 287), which may lead to a reduction in their novel ideas. Teachers and educators are encouraged to shift their methods of teaching away from traditional approaches to teaching to approaches such as REAPS in which they focus on developing creative ability among students.

In contrast, fluency, flexibility, and elaboration scores among students in upper grades increased significantly when compared with those of students in lower grades. For example, the pretest mean in elaboration for students in grade 4 during Year 1 and grade 5 during Year 2 was 7.1, while the posttest mean increased to 27.78; the pretest mean in elaboration for students who were in grade 5 in Year 1 and grade 6 in Year 2 was 3.86, while the posttest mean increased to 23.30. Only 3% of the total students in upper grades were with low implementers for two years and the rest of them were with high implementers for two years, high implementers in Year 1 and then with low implementers in Year 2, or low implementers in Year 1 and then with high implementers in Year 2. These findings may provide evidence that these abilities (fluency, flexibility, and elaboration) are linked more with developmental age and acquisition of knowledge than is originality, which is associated with the lower grades.

The results for Question 1 are similar to the results of previous studies of the REAPS model, which is an effective teaching model for helping students engage in the learning process, challenge their abilities, and develop their skills (Gomez-Arizaga et al., 2016; Riley et al., 2017;

Webber et al., 2018; Wu et al., 2015; Zimmerman et al., 2011). The REAPS model has been implemented in various countries, including the United States, China, Saudi Arabia, New Zealand, Australia, and Indonesia with students from middle or low socioeconomic levels. It had a positive effect on their abilities and motivation toward learning (Alhusaini, 2016; Alhusaini et al., 2015; Gomez-Arizaga et al.; Maker et al., 2015; Riley et al.; Reinoso, 2011; Webber et al.; Wu et al.; Yulindar et al., 2018; Zimmerman et al.). The results of the current study provided support for the claims from the developers of REAPS that this model has a positive influence on students' abilities, because such abilities in general have improved significantly from pretest to posttest (Maker & Zimmerman, 2008; Maker et al., 2015).

Discussion of Results for Research Question 2

The purpose of this question was to examine the influence of levels of FOI of the REAPS model on students' growth in creative problem solving in science. In general, differences were found among the levels of FOI in two aspects of creativity: flexibility and elaboration. Students who were with high implementers for two consecutive years showed greater increases in their elaboration ability than students who were with low implementers for two consecutive years. The mean for pretest scores in elaboration for students who were with high implementers for two consecutive years was 7.25, while the posttest mean for these students increased to 19.42; for students who were with low implementers for two consecutive years, the pretest mean was 6.09, and the posttest mean was 11.44. Sharp increases in means were found for students in classrooms of high implementers.

These findings are similar to the results of other studies in which teachers' levels of implementation affected students' growth in creativity. For example, Maker et al. (1996) examined the influence of the use of the DISCOVER curriculum model on students' problem solving ability as measured by the DISCOVER assessments. The researchers implemented their research in two classrooms (high and middle implementers). They found that students in a classroom with a high implementer increased in their creative problem solving ability in spatial and mathematics domains when compared with students with a middle implementer.

Other example is a study by Maker et al. (2008) in which they investigated the impact of implementation of the DISCOVER curriculum model on students' general creativity as measured by the Test of Creative Thinking–Drawing Production. The study took place over a three-year period and the researchers divided students into five groups based on levels of implementation: students who were with high implementers, students who were with middle implementers, students who were with low implementers, students who were with high implementers then moved to low implementers' classrooms, and students who were with low implementers and then moved to classrooms with high implementers. They found that the growth in creativity was associated with the levels of implementation and the period of exposure to the intervention. Students who were with high implementers in Years 1 and 2, middle implementers in Years 1 and 2, or low implementers in Year 1 and high implementers in Year 2 showed statistically significant growth compared with students who were with low implementers in Years 1 and 2 or with high implementers in Year 1 and low implementers in Year 2. However, in Year 3, only students who were with low implementers then moved to classrooms with high implementers showed statistically significant growth compared with the other four groups. Jo and Maker (2011) studied the effect of the use of the DISCOVER curriculum model on students' creativity in mathematics for one year. The researchers found that students' creativity in mathematics increased among students in classrooms with high or middle implementers compared with students in classrooms with low implementers. In addition, Maker et al. (2006) found that students whose teachers implemented the DISCOVER curriculum model at a high or middle level showed increases in their levels of creativity compared to students whose teachers implemented the DISCOVER curriculum model at a low level. However, no differences were found in general creativity of students in Year 3 among the three levels of implementation.

The growth of creative problem solving in science among students of high implementers compared with students of low implementers may be attributed to several factors. First, the high implementer teachers may create a safer and healthier environment for their students to express their ideas and expand their thoughts (elaboration) than low implementers. Thus, providing professional development workshops to help teachers know how to create an environment where students can have a space and feel free to exhibit their abilities and develop their skills is highly recommended. Second, approximately 93% of students who were with a high implementer for two years consisted of students from upper grades while approximately 78% of students who were with low implementers for two years came from lower grades, so the developmental age may have contributed to finding the statistically significant differences between these two groups. Third, high implementers may help students understand and apply scientific principles to solving local, personally-relevant problems and provide more details, which would develop growth in elaboration.

Other findings for Question 2 were that students who were with low implementers in Year 1 and with high implementers in Year 2 showed greater increases in their flexible thinking than other groups. Most of the students (80.4%) in this group were in grade 3 in Year 1 and moved to grade 4 in Year 2. The mean of their scores in flexibility on the pretest was 32.71, which increased to 43.00 on the posttest. In contrast, the means of scores of other groups increased only slightly. For example, the mean scores of students who were with high implementers in Year 1 and with low implementers in Year 2 on the pretest was 31.96, which increased to 35.06 on the posttest. The increase in flexibility of thinking among students in this group may be attributed to high implementers in Year 2 encouraging students to think differently and to freely examine situations from different angles. The results matched a conclusion of researchers in another study in which they found that students who moved from low implementers to high implementers demonstrated greater growth in their creativity than those who were with high implementers in Years 2 and 3, students who were with middle implementers in Years 2 and 3, students who were with low implementers in Years 2 and 3, and students who were with high implementers in Year 2 then moved to low implementers' classrooms in Year 3 (Maker et al., 2008) and the researchers emphasized that "If children experience an environment in which their unique ways of thinking are valued, and then encounter a teacher who does not provide this kind of support, they may become wary of expressing themselves" (p. 415). When students become accustomed to the methods of a teacher at a high level of implementation then move to a classroom in which the teacher uses more traditional methods of teaching, their creativity may not continue to develop.

One important finding for Question 2 is that students who were with low implementers for two years did not show statistically significant growth when compared with other groups. These results may encourage researchers and educators to consider levels of FOI when they look at students' scores and associate these scores with teachers' abilities to deliver an intervention before giving their conclusions about students' abilities. Students may have potential and readiness in specific areas of knowledge and because they are in classrooms with low implementers, their scores may be affected negatively. As a result, we should consider the claim from other researchers and accept the idea that a high level of FOI is related to desirable outcomes while a low FOI is not usually connected with desirable results (Harn et al., 2013; Maker et al., 2006; Maker et al., 2008).

No statistically significant differences were found between growth in creative problem solving in science scores and levels of FOI in total scores for TCPS-S, fluency, and originality. The variations between teachers who were identified as high and low implementers may not have been large enough to result in significant differences in promoting creativity in science in their students. These minor variations might be attributed to a number of factors. First, teachers at each grade level designed teaching units as groups and they met once every month to develop and discuss their teaching plans. Second, experts in the implementation of the REAPS model followed up with both high and low implementers and provided feedback to teachers to better implement REAPS. Thus, low implementers might have benefited from the feedback from these experts to increase their fidelity of implementation. Third, all teachers received support from school administrators and parents to help them implement the REAPS model in their classrooms. A curriculum supervisor who was a high implementer and familiar with REAPS visited classrooms, assessed teachers, and helped them to improve their fidelity of implementation of REAPS, especially if they were low implementers or those who were new to the school and the REAPS model. Finally, high implementers, who attended professional development workshops, mentored and taught other teachers so the quality of teaching of the low implementers might have been affected positively in the second year.

Discussion of Results for Research Question 3

The purpose of this question was to explore factors of the TCPS-S that were affected by the level of fidelity of implementation of the teachers. In general, students' scores across all levels of FOI increased from pretests to posttests in all three factors of the TCPS-S: Finding

Problems, Generating Detailed Solutions, and Classifying Elements. The increases in students' scientific creativity provide evidence that the REAPS model can be implemented in any educational environment and in different cultural contexts (Maker & Zimmerman, 2008, Maker et al., 2015). In the current study, more than half of the students in the school spoke languages other than English, and they represented different ethnic groups and socioeconomic levels.

Students who were with high implementers for two consecutive years, or with high implementers in the second year only, had statistically significant increases in their abilities for two factors, Generating Detailed Solutions and Classifying Elements, when compared with students who were with low implementers for two consecutive years or with low implementers in the second year. However, no differences were found between the four levels of FOI on the Finding Problems factor.

Although the differences between the four groups of FOI on the Generating Detailed Solutions factor were not distinguished by post hoc tests, the growth of scores of students who were with low implementers in Year 1 and with high implementers in Year 2 had a large increase from pretest (13.37) to posttest (19.95) when compared to the other groups. These findings provide evidence that high implementers helped students to produce creative solutions to problems more than did the low implementers.

For Classifying Elements, students who were with high implementers in both years or students who were with low implementers in Year 1 and with high implementers in Year 2 showed increases in their classification abilities. These findings are similar to the results of other studies (Jo & Maker, 2011; Maker et al., 1996; Maker et al., 2008). For example, Jo and Maker (2011) examined the effect of the teachers' implementation of the DISCOVER curriculum model on students' knowledge and creativity in mathematics. The DISCOVER curriculum was implemented for one year with students in grades 1 through 5. The findings were statistically significant for students in grades 2 and 3. Students with high implementers in grade 2 had higher scores in creativity in mathematics compared to students with middle implementer teachers. All results in grade 3 were statistically significant among students in classrooms with high, middle, and low implementers in both creativity in mathematics and content knowledge. For students in grades 1, 4, and 5, the results were not statistically significant in both creativity in mathematics and content knowledge across the three levels of implementation. One difference between the findings of Jo and Maker and the current study is that Jo and Maker examined the influence of levels of FOI at each grade level separately while the researcher in the present study combined all students from different grade levels based in their teachers' levels of FOI.

Maker et al. (1996) found that students in a classroom with a high implementer exhibited growth in creative problem solving ability in spatial and mathematics domains when compared with students with a middle implementer. Another example is a study by Maker et al. (2008) in which they investigated the influence of use of the DISCOVER curriculum model on students' general creativity for three years. Students were divided into five groups based on their levels of implementation: students who were with high implementers, students who were with middle implementers, students who were with low implementers, students who were with high implementers then moved to low implementers' classrooms, and students who were with low implementers and then moved to classrooms with high implementers. The researchers concluded that the period of exposure to the intervention (the DISCOVER curriculum model) affected the growth of students' creativity. If students participated in the intervention for two years with high implementers in Years 1 and 2, middle implementers in Years 1 and 2, or low implementers in Year 1 and high implementers in Year 2, their creativity increased significantly compared with students who were with low implementers in Years 1 and 2 or with high implementers in Year 1 and low implementers in Year 2. However, if the students were exposed to the intervention for Year 3, only students who were with low implementers then moved to classrooms with high implementers showed statistically significant growth compared with the other four groups. The assumption that a high level of FOI would lead to high or desirable outcomes while a low level of FOI would produce less desirable results (Harn, Parisi, & Stoolmiller, 2013; Maker et al., 2006; Maker et al., 2008; Sak & Maker, 2006) was supported in the current study and other studies.

Similar to the results for Question 2, students who were with low implementers for two years did not show statistically significant growth in any factors of the TCPS-S when compared with other groups. This conclusion gives scientific evidence for researchers to pay more attention to assess teachers' levels of FOI before making any final decision about the students' skills and abilities. Many students who are gifted or creative may not be recognized because they are in classrooms with teachers (low FOI) who do not use methods appropriate for their development (Harn et al., 2013; Maker et al., 2006; Maker et al., 2008).

Implications of the Study

REAPS is a promising teaching model that can be implemented to promote creativity in science among students. In studies that were conducted to examine the effect of REAPS, students stated that this model was a means to help them engage more in the learning process, challenge their abilities, and develop their interpersonal skills (Gomez-Arizaga et al., 2016; Riley et al., 2017; Webber et al., 2018; Wu et al., 2015). Likewise, teachers found REAPS to be an effective model that helped them to improve their teaching (Riley et al.; Zimmerman et al., 2011).

Practical Contribution

The results of this study will contribute to the improvement of practices in classrooms to serve all students for several reasons. First, the REAPS model has been applied in many classrooms since 2008, and educators and policymakers now have scientific evidence of the effectiveness of REAPS in the current study and previous studies (Alhusaini, 2016; Gomez-Arizaga et al., 2016; Riley et al., 2017; Webber et al., 2018; Wu et al., 2015; Zimmerman et al., 2011). Administrators can expand their adoption of the REAPS model in elementary schools in the United States and other countries to prepare young students to be creative problem solvers in science or other disciplines so they can face global challenges and serve their communities in the future.

Second, teacher educators can benefit from the results of this study by introducing REAPS to pre-service teachers so they are familiar with one of the teaching models that can help promote creativity in science among students. Pre-service teachers can benefit from the REAPS model and in-service teachers can implement it in their classrooms after participating in professional development workshops.

Third, the REAPS model was implemented in classrooms with students from different backgrounds and ethnicities as demonstrated in the current study. Thus, it can be adopted in any educational setting, as the developers of the REAPS model have claimed. Educators and researchers are strongly urged to apply this model in different school environments with students of different ethnicities and to examine its effect on students' creativity in science and in other disciplines.

Finally, the Test of Creative Problem Solving in Science (TCPS-S) was used in this study as an instrument to measure students' creativity in science. Most tests of creative problem solving are not content based, enabling teachers to measure students' creative application of the concepts and information they have learned in specific academic areas. In contrast, the TCPS-S is developed to measure both growth in content understating and growth in creative problem solving. Thus, teachers can use the TCPS-S to measure their students' creativity in science and then develop appropriate intervention programs to serve all students, especially these who are highly able.

Theoretical Contribution

Three contributions are associated with this study. First, four components of creativity are connected with developmental age: fluency, flexibility, elaboration, and originality. Students in upper grades in elementary school (i.e., grade 4 to 6) showed greater growth in fluency, flexibility, and elaboration scores than students in the lower grades. Thus, to develop these abilities, researchers and teachers are urged to consider introducing more learning experiences to develop these abilities in upper grade students. In contrast, originality is associated with the lower grades, so teachers can encourage original ideas by accepting them without judgment. The claim that novel ideas are negatively affected by formal education (Kim, 2011; Torrance, 1977; Urban, 1991) is supported by the results of the current study.

Second, students in classrooms of high implementers are more likely to experience greater increases in their abilities and skills than students in low implementers' classrooms. As a result, assessing teachers' levels of implementation is important for researchers to give clear explanations for a program's success or failure and to increase confidence that these outcomes came from the effectiveness of the program being evaluated (Nelson et al., 2012). When a program that is implemented with a high level of fidelity fails to produce the anticipated outcomes, researchers should reconsider the program design (Nelson et al.).

Third, perhaps an even more important theoretical implementation is that a model that was designed based on principles for teaching gifted children, such as REAPS, can be effective in developing scientific creativity in all children regardless of their levels of abilities (Alhusaini, 2016; Webber et al., 2018). Teachers who implement the REAPS model use the principles of differentiated instruction: content, process, product, and learning environment to provide different learning opportunities for all students to meet their abilities and interests.

Limitations of the Study

Five limitations should be considered. First, the instrument in this study (TCPS-S) was developed, administered, and scored in the English language and task completion is dependent on English ability. Thus, ESL students might not be able to demonstrate their creativity in science adequately. As a result, poor performance may result and this might affect their test scores.

Second, the study was a quasi-experimental design in which students were not randomly chosen to participate. In addition, no control group was included because the REAPS model was implemented school-wide. Thus, because of the sampling strategy, the results of this study should be interpreted with caution. This limitation was the major factor considered when designing a study of FOI rather than having experimental and control groups.

Third, student characteristics such as gender, ethnicity, socioeconomic level, and abilities were not reported in this study due to the school's policy. As a result, the author was not able to associate the impact of the REAPS model on students' creativity in science based on these characteristics.

Fourth, teachers designed their teaching units in groups based on the REAPS model principles. At each grade level, teachers developed joint teaching units and then implemented these units in their classrooms. Even though teaching units were reviewed by the research team, differentiating each teacher's participation in the design of the joint teaching units was not possible. Because the teams consisted of teachers with varying expertise and all contained high level implementers, those with more expertise helped those who were learning, which contributed to teachers being at similar levels of FOI across the school. This practice also contributed to the inability to distinguish each teacher's application of the model based on a reviewed of the teaching units.

Finally, the Test of Creative Problem Solving in Science (TCPS-S) has only one form. Students took pretests and posttests using the same form. Consequently, some students may not have taken the posttest seriously, or they simply may have been bored with the test.

Future Research

Further research is needed to examine the effect of the implementation of the REAPS model on students' creative problem solving. Researchers are encouraged to examine the influence of the use of the REAPS model on students' creativity in science in different school settings, such as middle schools, high schools, and colleges. Until now, three studies have been implemented in high school to examine the effect of the REAPS model on students' problem-solving ability and learning engagement (Riley et al., 2017; Webber et al., 2018; Yulindar et al., 2018). No study has been implemented in the middle school setting. Other possible studies can be conducted to examine the impact of the REAPS model on students' creativity in science in different classroom environments and educational programs, such as special classrooms or programs for gifted and special needs students.

Researchers are encouraged to investigate factors that affect fidelity of implementation of the REAPS model. These factors may include teachers' characteristics, students' characteristics, teacher to student ratios, school administrative support, program duration, and school resources. Future study also is recommended to expand the levels of implementation of the REAPS model. A categorical technique was used in which teachers were divided into high and low levels of implementation; future researchers may include a middle level as the third level of implementation, or they can use a continuous technique in which FOI is determined by percentages based on the quality of teachers' implementation of the REAPS model.

In this study, the cut-off point to differentiate high and low implementers was 4.5 out of 6. Thus, future studies to reduce this cut-off point to 3 and then investigate the effects of levels of FOI on students' creativity in science is recommended. Another suggested study is to examine the influences of FOI on students' creativity for each grade level separately. In the current study, the researcher was not able to examine these influences because the simple size was not large enough. The study extended for two years and only students who took both a pretest and posttest were included in the analysis. As a result, all students at different grade levels were combined based on the levels of FOI.

For the use of the TCPS-S to evaluate students' creativity in science, researchers and teachers are encouraged to examine the students' abilities in five areas: total scores of the test, fluency, flexibility, elaboration, and originality. Differences in growth in fluency, flexibility, and elaboration were linked with upper grades while differences in growth in originality was associated with lower grades. One possible future study is to re-examine these findings.

Students must be fluent in English to perform well and complete the TCPS-S tasks to exhibit their creativity in science. Thus, ESL students or students with disabilities may have faced difficulty completing the test and had low scores although they could be creative in science. To address this problem, figural tests in which no writing ability is needed may be developed to help ESL students or students with disabilities to demonstrate their creativity in science. Another suggested solution is to interview ESL students in their first languages to record their responses to the TCPS-S tasks. Developing versions of the TCPS-S that reflect students' languages may be another choice to allow these students to show their creativity in science.

One strong aspect of the current study is the variety of methods of collecting data to determine the levels of FOI. The methods of measuring FOI are the most important aspects of implementation research. The research team used the most popular methods for measuring FOI: classroom observations, ratings by experts, interviews, and self-reports by implementers (Durlak & DuPre, 2008; Dusenbury et al., 2003; Mowbray et al., 2003). Future researchers are encouraged to follow the same procedures for gathering data to determine the levels of FOI of their interventions.

Finally, the purpose of this study was to examine the effect of the use of the REAPS model on students' creativity in science. Future studies are suggested to examine the impact of the use of the REAPS model on students' creativity in different content areas such as mathematics, writing, design, and technology.

Conclusion

REAPS is a promising teaching model that can be implemented in classrooms to promote creativity in science among students. In general, students benefited from this model during two years of implementation. The originality scores of lower grade students increased, and fluency, flexibility, and elaboration increased among upper grade students. Two abilities were associated with high implementers, flexibility and elaboration, while the total scores in the TCPS-S, fluency, and originality were not significant at different levels of implementation. Students who were with high implementers for two consecutive years or with high implementers in the second year showed increases in two factors of the TCPS-S: Generating Detailed Solutions and Classifying Elements. However, no differences were found in the Finding Problems factor at different levels of FOI.

During the implementation of the REAPS model and after analyzing the TCPS-S scores, students met the common characteristics of creativity that various researchers discussed or proposed (Amabile, 1983; 1996; 2013; Csikszentmihalyi, 2006; Guilford, 1950; Hu & Adey, 2002; Isaksen et al, 2011; Maker; 1993; Mednick, 1962; Rhodes,1961; Torrance, 1988, 1993; Welsch, 1980; Zeng, Proctor, & Salvendy, 2009). First, students followed a series of processes to arrive at creative solutions. For example, one aspect of REAPS is Thinking Actively in a Social Context (TASC) in which students followed eight steps to study a real-world problem. Second, students used specific characteristics possessed by creative persons, such as imagination and openness to generate creative solutions or products. Third, some students produced creative outcomes that were novel, appropriate, and useful for their developmental stages. In the TCPS-S, the raters followed scientific procedures to determine original ideas (Maker et al., 2017) and used a statistical approach to distinguish between original and common ideas by sorting all ideas based on their frequencies among participants and then selecting ideas from the top ten percent. Finally, students were given opportunities to interact with their environments to generate unique outcomes. When implementing the REAPS model, teachers select real-world problems from the local area. In addition, when students took the TCPS-S, the problems that were presented to them were from the local natural environment in Sydney so students could relate to these problems. The teaching model being evaluated and the assessment tool were consistent.

One interesting note from the findings of the current study is that students who were with low implementers for two years did not show statistically significant growth when compared with other groups. Thus, we can accept the idea that a high level of FOI is associated with desirable outcomes while a low FOI is not linked with desirable results over time.

The REAPS model was developed to serve gifted students. However, non-gifted students can benefit from this model to improve their skills because the heart of this model depends on solving local problems that can be seen and felt by all students. The principles of differentiated instruction are integrated in the REAPS model to help teachers to meet students' abilities and interests. Students in the current study had varied levels of abilities. However, they showed growth in their creativity in science from pretest to posttest after being exposed to this model for two years.

APPENDIX A

A DESCRIPTIVE CLASSIFICATION OF STUDIES OF PBL, TASC, DISCOVER, AND REAPS

Classification of Studies by Research Design

Research Design	PBL		TASC		DISCOVER		REAPS		Total	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%
Qualitative							3*	37.5%	3	6.8%
Case study			1	25%			1	12.5%	2	4.6%
Descriptive	1	3.85%							1	2.3%
Action Research	2	7.7%	1	25%			1	12.5%	4	9.1%
Total	**3**	**11.55%**	**2**	**50%**	**0**	**0%**	**5**	**62.5%**	**10**	**22.8%**
Quantitative										
Quasi-experimental	17	65.4%	1	25%	6	100%	3	37.5%	27	61.3%
Action Research	3	11.5%							3	6.8%
Total	**20**	**76.9%**	**1**	**25%**	**6**	**100%**	**3**	**37.5%**	**30**	**68.1%**
Mixed Method	2	7.7%	1	25%					3	6.8%
Action Research	1	3.85%							1	2.3%
Total	**3**	**11.55%**	**1**	**25%**	**0**	**0%**	**0**	**0%**	**4**	**9.1%**
Total for Each Teaching Model	**26**	**100%**	**4**	**100%**	**6**	**100%**	**8**	**100%**	**44**	**100%**

*Researchers in these two studies did not specify the research designs.

Classification of Studies by Data Collection Instrument

Instrument Type	PBL		TASC		DISCOVER		REAPS		Total	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%
Interview	6	9%	2	22.2%	5	25%	3	23.1%	16	14.7%
Questionnaire	14	20.9%	3	33.4%					17	15.6%
Observation	7	10.4%	2	22.2%	5	25%	1	7.7%	15	13.8%
Survey	9	13.4%					1	7.7%	10	9.15%
Self-checklists	6	9%	1	11.1%	2	10%			9	8.3%
Study materials	2	3%	1	11.1%	1	5%	4	30.8%	8	7.3%
Video recordings	1	1.5%							1	0.9%
Test Scores	9	13.4%			1	5%	2	15.3%	12	11%
Standardized Test	3	4.5%			1	5%			4	3.7%
Creativity test (general domain)	4	5.9%			2	10%	1	7.7%	7	6.4%
Creativity test (specific domain)	6	9%			3	15%	1	7.7%	10	9.15%
Total for Each Teaching Model	**67**	**100%**	**9**	**100%**	**20**	**100%**	**13**	**100%**	**109**	**100%**

Classification of Studies by Setting

Element	PBL		TASC		DISCOVER		REAPS		Total	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%
School Level										
Kindergarten	3	11.5%			2	33.3%			5	11.5%
Elementary school	3	11.5%	3	75%	4	66.7%	5	62.5%	15	34.1%
Middle school	9	34.6%	1	25%					10	22.6%
High school	11	42.4%					3	37.5%	14	31.8%
Total	**26**	**100%**	**4**	**100%**	**6**	**100%**	**8**	**100%**	**44**	**100%**
School Area										
Urban	21	80.8%			2	33.3%	6	75.5%	29	65.9%
Suburban or Rural	1	3.8%					2	25.%	3	6.8%
Unknown	4	15.4%	4	100%	4	66.7%			12	27.3%
Total	**26**	**100%**	**4**	**100%**	**6**	**100%**	**8**	**100%**	**44**	**100%**
Classroom Size										
Fewer than 20 students	2	7.7%	1	25%	1	16.7%			4	9.1%
From 20 to 29 students	6	23.1%			1	16.7%	2	25.%	9	20.4%
30 students or more	13	50%	2	50%			1	12.5%	16	36.4%
Classroom size not stated	5	19.2%	1	25%	4	66.6	5	62.5%	15	34.1%
Total	**26**	**100%**	**4**	**100%**	**6**	**100%**	**8**	**100%**	**44**	**100%**
Classroom type										
Regular students	2	7.7%			3	50%	2	25%	7	15.9%
Gifted students	5	19.2%	1	25%	1	16.7%	2	25%	9	20.5%
Classroom type not stated	19	73.1%	3	75%	2	33.3%	4	50%	28	63.6%
Total	**26**	**100%**	**4**	**100%**	**6**	**100%**	**8**	**100%**	**44**	**100%**

Classification of Studies by Teacher Information

Element	PBL		TASC		DISCOVER		REAPS		Total	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%
Teacher(s) attended professional development workshops related to teaching model										
Yes	11	42.3%			5	83.3%	5	62.5%	21	47.7%
No information stated	15	57.7%	4	100%	1	16.7%	3	37.5%	23	52.3%
Total	**26**	**100%**	**4**	**100%**	**6**	**100%**	**8**	**100%**	**44**	**100%**
Teaching experience										
From 1 to 5 years	2									
From 5 to 10 years	3									
More than 10 years	4						1	12.5%		
No information stated	19	73.1%	4	100%	6	100%	7	87.5%	36	81.4%
Gender										
Female	2		1						3	6.8%
Male	3						1	12.5%	4	9.1%
No information stated	21		3		6		7	87.5%	37	84.1%
Total	**26**		**4**	100%	**6**	100%	**8**	100%	**44**	**100%**
Subject taught										
Science	25**		1		2**		8			
Mathematics	2**		1		5**					
Other subject(s)	2**		2		2**					

*In one study (Gallagher & Gallagher, 2013), teaching experience ranged from one to more than 10 years.

**Researchers in six studies examined the effect of the use of PBL on students' creativity in more than one subject. Anazifa and Djukri (2017), Kuo et al. (2011), and Maker et al., (2008) included science and mathematics in their studies, Gallagher and Gallagher (2013) included science and social studies in their research, Gallagher and Stepien (1992) included science in their study, and Maker et al. (1996) included math and linguistics in their study.

The Strengths and Weakness Across Studies

Element	PBL		TASC		DISCOVER		REAPS		Total	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%
School setting was described										
Yes	10	38.5%			4	66.7%	7	87.5%	21	47.7%
No	16	61.5%	4	100%	2	33.3%	1	12.5%	23	52.3%
Teachers' demographic characteristics were described										
Yes	7	26.9%			1	16.7%			8	18.2%
No	19	73.1%	4	100%	5	83.3%	8	100%	36	81.8%
Students' demographic characteristics were described										
Yes	16	61.5%	3	75%	6	100%	5	62.5%	30	68.2%
No	10	38.5%	1	25%			3	37.5%	14	31.8%
Teachers were trained before implementing the teaching model										
Yes	11	42.3%			5	83.3%	4	50%	20	45.5%
No (Unknown)	15	57.7%	4	100%	1	16.7%	4	50%	24	54.5%
Content of workshop was described										
Yes	4	15.4%			4	66.7%	4	50%	12	27.3%
No (Unknown)	22	84.6%	4	100%	2	33.3%	4	50%	32	72.7%
Period of workshop was stated										
Yes	5	19.2%			4	66.7%	4	50%	13	29.5%
No (Unknown)	21	80.8%	4	100%	2	33.3%	4	50%	31	70.5%
Teachers continued receiving workshops and support during implementing the teaching model										
Yes	4	15.4%			1	16.7%	3	37.5%	8	18.2%
No (Unknown)	22	84.6%	4	100%	5	83.3%	5	62.5%	36	81.8%

(Continued)

The Strengths and Weakness Across Studies

Element	PBL		TASC		DISCOVER		REAPS		Total	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%
Test scores were included										
Yes	9	34.6%			6	100%	1	12.5%	16	36.4%
No	17	65.4%	4	100%			7	87.5%	28	63.6%
Standardized test was included										
Yes	8	30.8%	1	25%	4	66.7%	1	12.5%	14	31.8%
No	18	69.2%	3	75%	2	33.3%	7	87.5%	30	68.2%
Pretest and posttest were used										
Yes	18	69.2%	1	25%	1	16.7%	3	37.5%	23	52.3%
No	8	30.8%	3	75%	5	83.3%	5	62.5%	21	47.7%
Time between implementing the teaching model and collecting the data was stated										
Yes	22	84.6%	1	25%	5	83.3%	2	25%	30	68.2%
No	4	15.4%	3	75%	1	16.7%	6	75%	14	31.8%
Triangulation of data was used										
Yes	20	76.9%	3	75%	4	66.7%	3	37.5%	30	68.2%
No	6	23.1%	1	25%	2	33.3%	5	62.5%	14	31.8%
Fidelity of implementing the teaching model was measured										
Yes	6	23.1%			6	100%	3	37.5%	15	34.1%
No (Unknown)	20	76.9%	4	100%			5	62.5%	29	65.9%
Time of observing classroom was stated										
Yes	2	7.7%	1	25%	5	83.3%	3	37.5%	11	25%
No (Unknown)	24	92.3%	3	75%	1	16.7%	5	62.5%	33	75%

(Continued)

The Strengths and Weakness Across Studies

Element	PBL		TASC		DISCOVER		REAPS		Total	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%
Someone other than the researcher(s) observed classroom										
Yes	7	26.9%	1	25%	5	83.3%	1	12.5%	14	31.8%
No (Unknown)	19	71.1%	3	75%	1	16.7%	7	87.5%	30	68.2%
Practices of the teachers in control group were described										
Yes	13	50%			6	100%			19	43.2%
No (Unknown)	4	15.4%	1	25%			1	12.5%	6	13.6%
Not Apply	9	34.6%	3	75%			7	87.5%	19	43.2%
The results were statistically significant										
Yes	25	96.2%	1	25%	5	83.3%	6	75%	37	84.1%
No										
Not Apply	1	3.8%	3	75%	1	16.7%	2	25%	7	15.9%
The effect size was reported										
Yes	8	30.8%			5	83.3%	1	12.5%	14	31.8%
No (Unknown)	15	57.7%	1	25%	1	16.7%	1	12.5%	18	40.9%
Not Apply	3	11.5%	3	75%			6	75%	12	27.3%
Researcher(s) relationship in the study was stated										
Yes	13	50%	1	25%	3	50%	6	75%	23	52.3%
No	13	50%	3	75%	3	50%	2	25%	21	47.7%
The researcher(s) was the instructor for the treatment group										
Yes	6	23.1%	1	25%			1	12.5%	8	18.2%
No	20	76.9%	3	75%	6	100%	7	87.5%	36	81.8%

(Continued)

The Strengths and Weakness Across Studies

Element	PBL		TASC		DISCOVER		REAPS		Total	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%
The duration of the study was stated										
Yes	21	80.8%	2	50%	6	100%	6	75%	35	79.5%
No	5	19.2%	2	50%			2	25%	9	20.5%

APPENDIX B

RESULTS OF THE EVALUATION OF QUALITY OF PROBLEM-BASED LEARNING (PBL) STUDIES

(QUANTITATIVE, QUALITATIVE, AND MIXED-METHODS RESEARCH)

The Quantitative Research Design Studies of PBL (1-4)

Indicators	Aidoo et al., 2016	Anazifa & Djukri, 2017	Araz & Sungur, 2007	Dods, 1997	Drake & Long, 2009	Gallagher & Gallagher, 2013
1. Research Conceptualization						
1.1. Rationale of study was stated	5	5	5	5	5	5
1.2. Purpose of study was stated	5	5	5	5	5	5
1.3. Theoretical framework was discussed	3	3	1	N/A	5	3
1.4. Research questions were stated	1	1	5	5	5	5
2. Settings and Participants						
2.1. Description of the setting was provided	2	2	5	1	5	5
2.2. Description of how the setting was selected was provided	1	1	5	5	1	5
2.3. Procedures for selecting the participants were discussed	5	5	5	N/A	5	5
2.4. Relevant demographic information about participants was provided	2	2	5	5	5	5
2.5. Description of the role of researcher(s) in relation to the setting and participants was clear	1	1	1	5	5	5
3. Intervention						
3.1. Relevant demographic information of interventionists was provided	1	1	1	3	3	5
3.2. Interventionists' roles were described	5	5	5	5	5	5
3.3. Interventionists were equivalent across conditions	5	1	5	N/A	5	N/A
3.4. Description of the practice was stated clearly	5	5	5	5	5	5
3.5. Description of how the practice was implemented was included	3	5	5	3	5	5

(Continued)

The Quantitative Research Design Studies of PBL (1-4)

Indicators	Studies					
	Aidoo et al., 2016	Anazifa & Djukri, 2017	Araz & Sungur, 2007	Dods, 1997	Drake & Long, 2009	Gallagher & Gallagher, 2013
3.6. Relevant aspects of the practice in the control group were described	5	1	5	N/A	5	N/A
4. Fidelity of Implementation (FOI)						
4.1. FOI was assessed	1	1	1	1	1	1
4.2. Key features of the practice were assessed	1	1	1	1	1	1
4.3. Inter-observer reliability of FOI was collected regularly	1	1	1	1	5	1
5. Attrition						
5.1. Overall attrition was low across groups	5	5	5	5	5	5
5.2. Difference in percentage of attrition between groups was low	5	5	5	N/A	5	5
6. Measurements						
6.1. Multiple outcome measures were used	1	5	5	5	5	5
6.2. Data collection was described	5	3	1	5	3	5
6.3. At least one measure of generalized performance was used	5	1	5	1	1	5
6.4. Outcomes were measured at both pretest and posttest, at a minimum	5	5	1	5	5	1
6.5. Outcome measures had evidence of internal consistency	1	1	5	1	1	1
6.6. Inter-observer reliability was assessed and acceptable	N/A	1	N/A	1	N/A	5
6.7. Evidence of concurrent, content, construct, and/or predictive validity was provided.	1	1	5	1	1	1

(Continued)

The Quantitative Research Design Studies of PBL (1-4)

Indicators	Aidoo et al., 2016	Anazifa & Djukri, 2017	Araz & Sungur, 2007	Dods, 1997	Drake & Long, 2009	Gallagher & Gallagher, 2013
7. Data Analysis						
7.1. Data analysis technique was appropriate.	5	5	5	5	5	5
7.2. Effect size was reported.	1	1	5	5	5	5
Total Score	**86**	**79**	**108**	**84**	**112**	**109**
Total Percent	**61.4%**	**54.5%**	**77.1%**	**70%**	**80%**	**80.7%**

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008).
Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.
N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

The Quantitative Research Design Studies of PBL (2-4)

Indicators	Gallagher & Stepien, 1992	Gordon et al., 2001	Horak & Galluzzo, 2017	Inel & Balim, 2010	Jo & Ku, 2011	Mundilarto, 2017
1. Research Conceptualization						
1.1. Rationale of study was stated	5	5	5	5	5	5
1.2. Purpose of study was stated	5	5	5	5	5	5
1.3. Theoretical framework was discussed	3	3	5	3	1	3
1.4. Research questions were stated	5	1	5	5	1	1
2. Settings and Participants						
2.1. Description of the setting was provided	5	5	5	2	3	2
2.2. Description of how the setting was selected was provided	5	5	5	1	1	1
2.3. Procedures for selecting participants were discussed	5	1	5	5	5	5
2.4. Relevant demographic information about participants was provided	1	5	5	2	2	2
2.5. Description of role of researcher(s) in relation to the settings and participants was clear	1	1	5	1	5	1
3. Intervention						
3.1. Relevant demographic information of interventionists was provided	1	1	5	1	1	1
3.2. Interventionists' roles were described	5	5	5	5	5	1
3.3. Interventionists were equivalent across conditions	5	5	5	5	N/A	5
3.4. Description of the practice was stated clearly	5	5	5	5	5	5
3.5. Description of how the practice was implemented was included	5	5	5	5	5	1

(Continued)

The Quantitative Research Design Studies of PBL (2-4)

Indicators	Gallagher & Stepien, 1992	Gordon et al., 2001	Horak & Galluzzo, 2017	Inel & Balim, 2010	Jo & Ku, 2011	Mundilarto, 2017
3.6. Relevant aspects of the practice in the control group were described	1	1	5	5	N/A	5
4. Fidelity of Implementation (FOI)						
4.1. FOI was assessed	1	1	5	1	1	3
4.2. Key features of practice were assessed	1	1	5	1	1	1
4.3. Inter-observer reliability of FOI was collected regularly	1	1	5	1	1	N/A
5. Attrition						
5.1. Overall attrition was low across groups	3	5	5	5	5	5
5.2. Difference in percentage of attrition between groups was low	3	5	5	5	5	5
6. Measurements						
6.1. Multiple outcome measures were used	1	5	5	5	5	5
6.2. Data collection was described	5	5	5	5	5	5
6.3. At least one measure of generalized performance was used	1	1	5	1	1	1
6.4. Outcomes were measured at both pretest and posttest, at a minimum	5	1	5	5	5	5
6.5. Outcome measures had evidence of internal consistency	1	1	3	1	1	1
6.6. Inter-observer reliability was assessed and acceptable	N/A	N/A	5	N/A	N/A	N/A
6.7. Evidence of concurrent, content, construct, and/or predictive validity was provided	1	1	1	5	1	1

(Continued)

The Quantitative Research Design Studies of PBL (2-4)

Indicators	Gallagher & Stepien, 1992	Gordon et al., 2001	Horak & Galluzzo, 2017	Inel & Balim, 2010	Jo & Ku, 2011	Mundilarto, 2017
7. Data Analysis						
7.1. Data analysis technique was appropriate	5	1	5	5	5	5
7.2. Effect size was reported	1	1	5	1	1	1
Total Scores	**86**	**82**	**139**	**96**	**81**	**81**
Total Percent	**61.4%**	**58.6%**	**95.9%**	**68.6%**	**62.3%**	**60%**

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008).
Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.
N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

The Quantitative Research Design Studies of PBL (3-4)

Indicators	Studies					
	Nurdin & Setiawan, 2015	Ratnasari et al., 2017	Siew & Chin, 2018	Siew et al., 2017	Siew & Mapeala, 2016	Sungur et al., 2006
1. Research Conceptualization						
1.1. Rationale of study was stated	5	5	5	5	5	5
1.2. Purpose of study was stated	5	5	5	5	5	5
1.3. Theoretical framework was discussed	1	1	5	5	5	1
1.4. Research questions were stated	1	1	5	5	5	5
2. Settings and Participants						
2.1. Description of the setting was provided	2	2	2	2	2	2
2.2. Description of how the setting was selected was provided	1	1	5	5	5	1
2.3. Procedures for selecting the participants were discussed	1	1	5	5	5	5
2.4. Relevant demographic information about participants was provided	2	2	5	5	5	5
2.5. Description of the role of researcher(s) in relation to the settings and participants was clear	5	1	5	5	5	1
3. Intervention						
3.1. Relevant demographic information of interventionists was provided	1	1	3	1	5	1
3.2. Interventionists' roles were described	5	5	5	5	5	5
3.3. Interventionists were equivalent across conditions	5	N/A	5	5	5	5
3.4. Description of the practice was stated clearly	5	1	5	5	5	5
3.5. Description of how the practice was implemented was included	5	3	5	5	5	5

(Continued)

The Quantitative Research Design Studies of PBL (3-4)

Indicators	Studies					
	Nurdin & Setiawan, 2015	Ratnasari et al., 2017	Siew & Chin, 2018	Siew et al., 2017	Siew & Mapeala, 2016	Sungur et al., 2006
3.6. Relevant aspects of the practice in the control group were described	5	N/A	5	5	5	5
4. Fidelity of Implementation (FOI)						
4.1. FOI was assessed	1	1	5	5	3	1
4.2. Key features of the practice were assessed	1	1	1	1	1	1
4.3. Inter-observer reliability of FOI was collected regularly	1	1	1	1	1	1
5. Attrition						
5.1. Overall attrition was low across groups	5	N/A	5	5	5	5
5.2. Difference in percentage of attrition between groups was low	5	N/A	5	5	5	5
6. Measurements						
6.1. Multiple outcome measures were used	5	5	1	1	1	5
6.2. Data collection was described	5	1	5	5	5	5
6.3. At least one measure of generalized performance was used	1	1	5	5	5	1
6.4. Outcomes were measured at both pretest and posttest, at a minimum	5	1	5	5	5	5
6.5. Outcome measures had evidence of internal consistency	1	1	5	5	1	1
6.6. Inter-observer reliability was assessed and acceptable	1	1	1	1	1	1
6.7. Evidence of concurrent, content, construct, and/or predictive validity was provided	1	1	5	5	3	3

(Continued)

The Quantitative Research Design Studies of PBL (3-4)

Indicators	Nurdin & Setiawan, 2015	Ratnasari et al., 2017	Siew & Chin, 2018	Siew et al., 2017	Siew & Mapeala, 2016	Sungur et al., 2006
7. Data Analysis						
7.1. Data analysis technique was appropriate	5	1	5	5	5	5
7.2. Effect size was reported	5	1	1	5	5	1
Total Scores	**91**	**45**	**120**	**122**	**118**	**96**
Total Percent	**62.8%**	**36%**	**82.8%**	**84.1%**	**81.4%**	**66.2%**

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008).
Scores given ranged from 1 to 5. 1= Inadequate/Unknown, 2= weakly stated, 3= Partial, 4= good, and 5= Adequate.
N/A= Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

The Quantitative Research Design Studies of PBL (4-4)

Indicators	Studies	
	Wartono et al., 2018	Wong & Day, 2009
1. Research Conceptualization		
1.1. Rationale of study was stated	5	5
1.2. Purpose of study was stated	5	5
1.3. Theoretical framework was discussed	1	5
1.4. Research questions were stated	1	5
2. Settings and Participants		
2.1. Description of the setting was provided	2	5
2.2. Description of how the setting was selected was provided	1	5
2.3. Procedures for selecting participants were discussed	3	5
2.4. Relevant demographic information about participants was provided	3	5
2.5. Description of the role of researcher(s) in relation to the settings and participants was clear	1	5
3. Intervention		
3.1. Relevant demographic information of interventionist(s) was provided	1	5
3.2. Interventionists' roles were described	5	5
3.3. Interventionists were equivalent across conditions	5	5
3.4. Description of the practice was stated clearly	3	5
3.5. Description of how the practice was implemented was included	5	5
3.6. Relevant aspects of the practice in the control group was described	1	5
4. Fidelity of Implementation (FOI)		
4.1. FOI was assessed	1	1
4.2. Key features of the practice were assessed	1	1
4.3. Inter-observer reliability of FOI was collected regularly	1	1

(Continued)

The Quantitative Research Design Studies of PBL (4-4)

	Studies	
5. Attrition		
5.1. Overall attrition was low across groups	5	5
5.2. Difference in percentage of attrition between groups was low	5	5
6. Measurements		
6.1. Multiple outcome measures were used	1	5
6.2. Data collection was described	1	5
6.3. At least one measure of generalized performance was used	5	1
6.4. Outcomes were measured at both pretest and posttest, at a minimum	1	5
6.5. Outcome measures had evidence of internal consistency	1	1
6.6. Inter-observer reliability was assessed and acceptable	1	1
6.7. Evidence of concurrent, content, construct, and/or predictive validity was provided	3	1
7. Data Analysis		
7.1. Data analysis technique was appropriate	5	5
7.2. Effect size was reported	1	1
Total Scores	**74**	**113**
Total Percent	**51 %**	**77.9%**

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008).
Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.
N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

The Qualitative Research Design Studies of PBL

Indicators	Studies		
	Haridza & Irving, 2017	Nurdyani et al., 2018	Zhang et al., 2011
1. Research Conceptualization			
1.1. Rationale of study was stated	5	5	5
1.2. Purpose of study was stated	5	5	5
1.3. Theoretical framework was discussed	5	1	1
1.4. Research questions were stated	5	1	5
2. Settings and Participants			
2.1. Description of the setting was provided	3	2	5
2.2. Description of how the setting was selected was provided	5	1	5
2.3. Procedures for selecting the participants were discussed	N/A	1	5
2.4. Relevant demographic information about participants was provided	5	2	5
2.5. Description of the role of researcher(s) in relation to the setting and participants was clear	5	1	5
3. Intervention			
3.1. Relevant demographic information of interventionists was provided	5	5	5
3.2. Interventionists' roles were described	5	1	5
3.3. Description of the practice was stated clearly	5	5	5
3.4. Description of how the practice was implemented was included	5	3	5
4. Data Sources and Analysis			
4.1. Data collection was described	3	2	5
4.2. Multiple data sources were included	5	5	5
4.3. Framework for data analysis was described	1	2	3
4.4. Coding procedures were described	1	1	1
4.5. Identification of themes was described	1	1	1

(Continued)

The Qualitative Research Design Studies of PBL

Indicators	**Studies**		
	Haridza & Irving, 2017	Nurdyani et al., 2018	Zhang et al., 2011
5. Trustworthiness and Credibility			
5.1. Trustworthiness and credibility were established	1	1	1
5.2. Audit trail was clear and detailed enough to follow	1	1	1
6. Outcome(s)			
6.1. Description of outcomes was clear	5	3	5
6.2. Strength of the outcomes or effects was stated	5	5	5
Total Scores	**81**	**54**	**88**
Total Percent	**77.1%**	**49.1.%**	**80%**

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008).
Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.
N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

The Mixed-Methods Design Studies of PBL

Indicators	Akınoğlu & Tandoğan, 2007	Fatimah, 2015	Siew et al., 2015
1. Research Conceptualization			
1.1. Rationale of study was stated	5	5	5
1.2. Purpose of study was stated	5	5	5
1.3. Theoretical framework was discussed	3	N/A	5
1.4. Research questions were stated	5	1	5
2. Settings and Participants			
2.1. Description of the setting was provided	2	3	5
2.2. Description of how the setting was selected was provided	5	1	5
2.3. Procedures for selecting participants were discussed	1	N/A	5
2.4. Relevant demographic information about participants was provided	3	3	5
2.5. Description of the role of researcher(s) in relation to the setting and participants was clear	5	1	5
3. Intervention			
3.1. Interventionists' roles were described	5	5	5
3.2. Demographic information of interventionists was provided	1	1	1
3.3. Interventionists were equivalent across conditions	5	N/A	N/A
3.4. Description of the practice was stated clearly	5	5	5
3.5. Description of how the practice was implemented was included	5	3	5
4. Fidelity of Implementation (FOI)			
4.1. FOI was assessed	1	1	1
4.2. Key features of the practice were assessed	1	1	1
4.3. Inter-observer reliability of FOI was collected regularly	1	1	1
5. Measurements			
5.1. Data collection was described	5	3	5
5.2. Multiple data sources were included	5	5	5

(Continued)

The Mixed-Methods Design Studies of PBL

Indicators	Studies		
	Akınoğlu & Tandoğan, 2007	Fatimah, 2015	Siew et al., 2015
5.3. Outcome measures had evidence of internal consistency	5	1	5
5.4. Inter-observer reliability was assessed and acceptable	5	1	N/A
5.5. Evidence of concurrent, content, construct, and/or predictive validity was provided	5	1	3
6. Data Analysis			
6.1. Data analysis technique was appropriate	5	1	5
6.2. Coding procedures were described	1	1	5
6.3. Identification of themes was described	1	1	5
7. Trustworthiness and Credibility			
7.1. Trustworthiness and credibility were established	1	1	5
7.2. Audit trail was clear and detailed enough to follow	1	1	5
8. Outcome(s)			
8.1. Description of outcomes was clear	5	5	5
8.2. Strength of the outcomes or effects was stated	5	5	5
8.3. At least one measure of generalized performance was used	5	1	5
8.4. Effect size was reported	1	1	1
Total Scores	**108**	**64**	**123**
Total Percent	**69.7%**	**45.7%**	**84.8%**

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008). In the CEC, only quality indicators for qualitative and quantitative research were developed. I combined the quality indicators of both qualitative and quantitative research for the mixed-methods research design.

Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.

N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percenta.

APPENDIX C

RESULTS OF THE EVALUATION OF QUALITY OF THE THINKING ACTIVELY IN A SOCIAL CONTEXT (TASC) STUDIES

(QUANTITATIVE, QUALITATIVE, AND MIXED-METHODS RESEARCH)

The Quantitative Research Design Studies of TASC

Indicators	Studies
	Abu Awwad et al., 2014
1. Research Conceptualization	
1.1. Rationale of study was stated	5
1.2. Purpose of study was stated	5
1.3. Theoretical framework was discussed	5
1.4. Research questions were stated	5
2. Settings and Participants	
2.1. Description of the setting was provided	1
2.2. Description of how the setting was selected was provided	1
2.3. Procedures for selecting participants were discussed	N/A
2.4. Relevant demographic information about participants was provided	2
2.5. Description of the role of researcher(s) in relation to the setting and participants was clear	1
3. Intervention	
3.1. Relevant demographic information of interventionists was provided	1
3.2. Interventionists' roles were described	1
3.3. Interventionists were equivalent across conditions	1
3.4. Description of the practice was stated clearly	5
3.5. Description of how the practice was implemented was included	2
3.6. Relevant aspects of the practice the in control group were described	1
4. Fidelity of Implementation (FOI)	
4.1. FOI was assessed	2
4.2. Key features of the practice were assessed	1
4.3. Inter-observer reliability of FOI was collected regularly	1
5. Attrition	
5.1. Overall attrition was low across groups	5
5.2. Difference in percentage of attrition between groups was low	5

(Continued)

The Quantitative Research Design Studies of TASC

Indicators	**Studies**
	Abu Awwad et al., 2014
6. Measurements	
6.1. Multiple outcome measures were used	5
6.2. Data collection was described	5
6.3. At least one measure of generalized performance was used	N/A
6.4. Outcomes were measured at both pretest and posttest, at a minimum	5
6.5. Outcome measures had evidence of internal consistency	5
6.6. Inter-observer reliability was assessed and acceptable	N/A
6.7. Evidence of concurrent, content, construct, and/or predictive validity was provided	5
7. Data Analysis	
7.1. Data analysis technique was appropriate	5
7.2. Effect size was reported	1
Total Scores	**81**
Total Percent	**62.3%**

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008).
Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.
N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

The Qualitative Research Design Studies of TASC

Indicators	Studies	
	Davies, 2008	Faulkner, 2008
1. Research Conceptualization		
1.1. Rationale of study was stated	5	5
1.2. Purpose of study was stated	5	5
1.3. Theoretical framework was discussed	5	5
1.4. Research questions were stated	2	3
2. Settings and Participants		
2.1. Description of the setting was provided	1	1
2.2. Description of how setting was selected was provided	2	1
2.3. Procedures for selecting participants were discussed	N/A	5
2.4. Relevant demographic information about participants was provided	3	5
2.5. Description of the role of researcher(s) in relation to the settings and participants was clear	5	1
3. Intervention		
3.1. Relevant demographic information of interventionists was provided	5	1
3.2. Interventionists' roles were described	1	1
3.3. Description of the practice was stated	2	5
3.4. Description of how the practice was implemented was included	3	3
4. Data Sources and Analysis		
4.1. Data collection was described	3	5
4.2. Multiple data sources were included	5	3
4.3. Framework for data analysis was described	1	5
4.4. Coding procedures were described.	1	N/A
4.5. Identification of themes was described	1	N/A
5. Trustworthiness and Credibility		
5.1. Trustworthiness and credibility were established	1	1
5.2. Audit trail was clear and detailed enough to follow	1	5

(Continued)

The Qualitative Research Design Studies of TASC

Indicators	Studies	
	Davies, 2008	Faulkner, 2008
6. Outcome(s)		
6.1. Description of outcomes was clear	3	3
6.2. Strength of the outcomes or effects was stated	3	5
Total Scores	**58**	**68**
Total Percent	**55.2%**	**68%**

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008).
Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.
N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

The Mixed-Methods Design Studies of TASC

Indicators	Studies
	West, 2008
1. Research Conceptualization	
1.1. Rationale of study was stated	1
1.2. Purpose of study was stated	5
1.3. Theoretical framework was discussed	5
1.4. Research questions were stated	1
2. Settings and Participants	
2.1. Description of the setting was provided	2
2.2. Description of how the setting was selected was provided	1
2.3. Procedures for selecting participants were discussed	1
2.4. Relevant demographic information about participants was provided	2
2.5. Description of the role of researcher(s) in relation to the setting and participants was clear	1
3. Intervention	
3.1. Interventionists' roles were described.	1
3.2. Relevant demographic information of interventionists was provided	1
3.3. Interventionists were equivalent across conditions	N/A
3.4. Description of the practice was stated clearly	1
3.5. Description of how the practice was implemented was included	1
4. Fidelity of Implementation (FOI)	
4.1. FOI was assessed	1
4.2. Key features of the practice were assessed	1
4.3. Inter-observer reliability of FOI was collected regularly	1
5. Measurements	
5.1. Data collection was described	3
5.2. Multiple data sources were included	5
5.3. Outcome measures had evidence of internal consistency	N/A
5.4. Inter-observer reliability was assessed and acceptable	N/A

(Continued)

The Mixed-Methods Design Studies of TASC

Indicators	Studies
	West, 2008
5.5. Evidence of concurrent, content, construct, and/or predictive validity was provided	N/A
6. Data Analysis	
6.1. Data analysis technique was appropriate	1
6.2. Coding procedures were described	1
6.3. Identification of themes was described	1
7. Trustworthiness and Credibility	
7.1. Trustworthiness and credibility were established	1
7.2. Audit trail was clear and detailed enough to follow	1
8. Outcome(s)	
8.1. Description of outcomes was clear	3
8.2. Strength of the outcomes or effects was stated	3
8.3. At least one measure of generalized performance was used	1
8.4. Effect size was reported	N/A
Total Scores	**46**
Total Percent	**35.4%**

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008). In the CEC, only quality indicators for qualitative and quantitative research were developed. I combined the quality indicators of both qualitative and quantitative research for the mixed-methods research design.

Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.

N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

APPENDIX D

RESULTS OF THE EVALUATION OF QUALITY OF THE DISCOVERING INTELLECTUAL STRENGTHS AND CAPABILITIES WHILE OBSERVING VARIED ETHNIC RESPONSES (DISCOVER) STUDIES (QUANTITATIVE RESEARCH)

The Quantitative Research Design Studies of DISCOVER

Indicators	Studies					
	Jo & Maker, 2011	Kuo et al., 2011	Maker et al., 2008	Maker et al., 2006	Maker et al., 1996	Sak & Maker, 2006
1. Research Conceptualization						
1.1. Rationale of study was stated	5	5	5	5	5	5
1.2. Purpose of study was stated	5	5	5	5	5	5
1.3. Theoretical framework was discussed	5	5	5	5	5	5
1.4. Research questions were stated	5	5	5	5	5	5
2. Settings and Participants						
2.1. Description of the setting was provided	3	3	3	3	5	3
2.2. Description of how the setting was selected was provided	5	5	5	5	5	5
2.3. Procedures for selecting participants were discussed	5	5	5	5	5	5
2.4. Relevant demographic information about participants was provided	5	5	5	5	5	5
2.5. Description of the role of researcher(s) in relation to the settings and participants was clear	5	5	1	1	5	1
3. Intervention						
3.1. Relevant demographics information of interventionists were provided	1	1	5	3	1	N/A
3.2. Interventionists' roles were described	5	5	5	5	5	N/A
3.3. Interventionists were equivalent across conditions	N/A	N/A	N/A	N/A	N/A	N/A

(Continued)

The Quantitative Research Design Studies of DISCOVER

Indicators	Studies					
	Jo & Maker, 2011	Kuo et al., 2011	Maker et al., 2008	Maker et al., 2006	Maker et al., 1996	Sak & Maker, 2006
3.4. Description of the practice was stated clearly	5	5	5	5	5	N/A
3.5. Description of how the practice was implemented was included	5	5	5	5	5	N/A
3.6. Relevant aspects of the practice in the control group were described	N/A	N/A	N/A	N/A	N/A	N/A
4. Fidelity of Implementation (FOI)						
4.1. FOI was assessed	5	5	5	5	5	N/A
4.2. Key features of the practice were assessed	5	5	5	5	5	N/A
4.3. Inter-observer reliability of FOI was collected regularly	5	5	5	5	5	N/A
5. Attrition						
5.1. Overall attrition was low across groups	N/A	N/A	N/A	N/A	N/A	N/A
5.2. Difference in percentage of attrition between groups was low	N/A	N/A	N/A	N/A	N/A	N/A
6. Measurements						
6.1. Multiple outcome measures were used	5	5	5	5	5	1
6.2. Data collection was described	5	5	5	5	5	5
6.3. At least one measure of generalized performance was used	5	1	5	5	1	5

(Continued)

The Quantitative Research Design Studies of DISCOVER

Indicators	Jo & Maker, 2011	Kuo et al., 2011	Maker et al., 2008	Maker et al., 2006	Maker et al., 1996	Sak & Maker, 2006
6.4. Outcomes were measured at both pretest and posttest, at a minimum	N/A	1	N/A	N/A	5	N/A
6.5. Outcome measures had evidence of internal consistency	1	N/A	5	5	1	N/A
6.6. Inter-observer reliability was assessed and acceptable	5	N/A	5	5	1	5
6.7. Evidence of concurrent, content, construct, and/or predictive validity was provided	1	N/A	5	5	1	5
7. Data Analysis						
7.1. Data analysis technique was appropriate	5	5	5	5	5	5
7.2. Effect size was reported	5	1	5	5	5	5
Total Scores	**106**	**92**	**114**	**112**	**105**	**70**
Total Percent	**88.3%**	**83.6%**	**95%**	**93.3%**	**84%**	**87.5%**

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008).
Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.
N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

APPENDIX E

RESULTS OF THE EVALUATION OF QUALITY OF THE REAL ENGAGEMENT IN ACTIVE PROBLEM SOLVING (REAPS) STUDIES

(QUANTITATIVE AND QUALITATIVE RESEARCH)

The Quantitative Research Design Studies of REAPS

Indicators	Studies		
	Alhusaini, 2016	Yulindar et al., 2018	Zimmerman et al., 2011
1. Research Conceptualization			
1.1. Rationale of study was stated	5	5	5
1.2. Purpose of study was stated	5	5	5
1.3. Theoretical framework was discussed	5	5	5
1.4. Research questions were stated	5	1	1
2. Settings and Participants			
2.1. Description of the setting was provided	5	2	3
2.2. Description of how the setting was selected was provided	5	1	1
2.3. Procedures for selecting participants were discussed	5	5	5
2.4. Relevant demographic information about participants was provided	3	1	2
2.5. Description of the role of researcher(s) in relation to the settings and participants was clear	5	1	5
3. Intervention			
3.1. Relevant demographic information of interventionists was provided	1	1	5
3.2. Interventionists' roles were described	5	1	5
3.3. Interventionists were equivalent across conditions	N/A	N/A	N/A
3.4. Description of the practice was stated clearly	5	5	5
3.5. Description of how the practice was implemented was included	5	1	5
3.6. Relevant aspects of the practice in the control group were described	N/A	N/A	N/A
4. Fidelity of Implementation (FOI)			
4.1. FOI was assessed	5	1	1
4.2. Key features of the practice were assessed	5	1	1
4.3. Inter-observer reliability of FOI was collected regularly	5	1	N/A
5. Attrition			
5.1. Overall attrition was low across groups	N/A	N/A	N/A

(Continued)

The Quantitative Research Design Studies of REAPS

Indicators	**Studies**		
	Alhusaini, 2016	Yulindar et al., 2018	Zimmerman et al., 2011
5.2. Difference in percentage of attrition between groups was low	N/A	N/A	N/A
. Measurements			
6.1. Multiple outcome measures were used	5	1	N/A
6.2. Data collection was described	5	5	5
6.3. At least one measure of generalized performance was used	5	1	5
6.4. Outcomes were measured at both pretest and posttest, at a minimum	5	5	5
6.5. Outcome measures had evidence of internal consistency	N/A	N/A	N/A
6.6. Inter-observer reliability was assessed and acceptable	5	N/A	N/A
6.7. Evidence of concurrent, content, construct, and/or predictive validity was provided	5	1	N/A
7. Data Analysis			
7.1. Data analysis technique was appropriate	5	3	5
7.2. Effect size was reported	5	N/A	1
Total Scores	**114**	**53**	**75**
Total Percent	**95%**	**48.2%**	**75%**

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008).

Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.

N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

The Qualitative Research Design Studies of REAPS

Indicators	Studies				
	Gomez-Arizaga et al., 2016	Riley et al., 2017	Reinoso, 2011	Webber et al., 2018	Wu et al,, 2015
1. Research Conceptualization					
1.1. Rationale of study was stated	5	5	1	5	5
1.2. Purpose of study was stated	5	5	3	5	5
1.3. Theoretical framework was discussed	5	3	3	3	5
1.4. Research questions were stated	5	5	1	5	5
2. Settings an d Participants					
2.1. Description of the setting was provided	5	5	5	5	5
2.2. Description of how the setting was selected was provided	5	5	5	5	5
2.3. Procedures for selecting the participants were discussed	5	1	N/A	1	5
2.4. Relevant demographics information about participants was provided	4	3	3	3	5
2.5. Description of the role of researcher(s) in relation to the settings and participants was clear	5	5	1	5	5
3. Intervention					
3.1. Relevant demographic information of interventionists was provided	2	1	3	1	3
3.2. Interventionists' roles were described	1	5	1	5	1
3.3. Description of the practice was stated clearly	5	5	5	5	5
3.4. Description of how the practice was implemented was included	2	5	5	3	3
4. Data Sources and Analysis					
4.1. Data collection was described	5	2	3	3	5
4.2. Multiple data sources were included	5	5	1	3	5

(Continued)

The Qualitative Research Design Studies of REAPS

Indicators	Studies				
	Gomez-Arizaga et al., 2016	Riley et al., 2017	Reinoso, 2011	Webber et al., 2018	Wu et al,, 2015
4.3. Framework for data analysis was described	5	1	1	1	5
4.4. Coding procedures were described	5	1	N/A	1	5
4.5. Identification of themes was described	5	1	N/A	1	5
5. Trustworthiness and Credibility					
5.1. Trustworthiness and credibility were established	5	1	N/A	1	5
5.2. Audit trail was clear and detailed enough to follow	5	1	N/A	1	5
6. Outcome(s)					
6.1. Description of outcomes were clear	5	5	3	5	5
6.2. Strength of the outcomes or effects was stated	5	5	3	5	5
Total Scores	**99**	**75**	**47**	**72**	**102**
Total Percent	**90%**	**68.2**	**55.3%**	**65.5%**	**92.7%**

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008).

Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.

N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

APPENDIX F

HIGHEST AND LOWEST SCORES RECEIVED BY STUDIES BASED ON RESEARCH DESIGN

Quantitative Research- Highest Scores Received

Author of Study	Horak & Galluzzo, 2017

Indicators	Scores	Justifications for Scores Given
1. Research Conceptualization		
1.1. Rationale of study was stated	5	The researchers discussed the importance of improving students' achievement in science using an active learning method.
1.2. Purpose of study was stated	5	The researchers investigated the impact of implementing PBL on students' academic achievement and classroom quality among students in middle schools.
1.3. Theoretical framework was discussed	5	The theoretical framework of this study was based on the constructivist learning theory.
1.4. Research questions were stated	5	One research question guided the inquiry.
2. Settings and Participants		
2.1. Description of the setting was provided	5	The researchers provided sufficient information about the setting of the study.
2.2. Description of how the setting was selected was provided	5	These schools for gifted students were identified by the school district to implement the PBL unit.
2.3. Procedures for selecting participants were discussed	5	The students were identified using an inclusive method for gifted students.
2.4. Relevant demographic information about participants was provided	5	The researchers provided sufficient information about the participants of the study.
2.5. Description of the role of researcher(s) in relation to the setting and participants was clear	5	The researchers stated that there were no potential conflicts of interest with respect to their research and that they did not receive any funding to conduct their study. They provided workshops to teachers in the treatment group.

(Continued)

Author of Study	Horak & Galluzzo, 2017

Indicators	Scores	Justifications for Scores Given
3. Intervention		
3.1. Relevant demographic information of interventionists was provided	5	Sufficient information about the teachers of the study was provided, such as their teaching experience, gender, and ethnicity.
3.2. Interventionists' roles were described	5	The teachers were facilitators for learning in the treatment group, while the teachers in control group used a direct teaching method in their classrooms.
3.3. Interventionists were equivalent across conditions	5	Three teachers were in each group; their characteristics were almost the same.
3.4. Description of the practice was stated clearly	5	The researchers described PBL and its components in the introduction of their study.
3.5. Description of how the practice was implemented was included	5	The researchers provided a detailed description of how PBL was implemented.
3.6. Relevant aspects of the practice in the control group were described	5	The students in the control group were taught with traditional teaching methods in which students were passive learners.
4. Fidelity of Implementation (FOI)		
4.1. FOI was assessed	5	Two researchers observed each teacher in the treatment group at the same time.
4.2. Key features of the practice were assessed	5	The researchers used a 5-point Likert-scale that consisted of 23 items, to gather information about FOI.
4.3. Inter-observer reliability of FOI was collected regularly	5	The inter-observer reliability ranged from 86% to 91%.
5. Attrition		
5.1. Overall attrition was low across groups	5	Attrition in the treatment group was about 9% and was about 5% in the control group. However, the researchers had a large sample size.
5.2. Difference in percentage of attrition between groups was low	5	The attrition among the two groups of students was low.

(Continued)

Author of Study	Horak & Galluzzo, 2017

Indicators	Scores	Justifications for Scores Given
6. Measurements		
6.1. Multiple outcome measures were used	5	Two instruments were used: performance on a standardized test in science and a self-rating survey.
6.2. Data collection was described	5	The researchers described in detail how they collected their data.
6.3. At least one measure of generalized performance was used	5	Performance on a standardized test in science is considered a measurement that can be used to generalize performance.
6.4. Outcomes were measured at both pretest and posttest, at a minimum	5	Both of the instruments were used to collect pretest and posttest data at the beginning and the end of the intervention.
6.5. Outcome measures had evidence of internal consistency	3	The researchers mentioned that the self-rating survey had a strong internal consistency. However, they did not report the value of it.
6.6. Inter-observer reliability was assessed and acceptable	5	The inter-observer reliability ranged from 86% to 91%.
6.7. Evidence of concurrent, content, construct, and/or predictive validity was provided	1	No information was stated.
7. Data Analysis		
7.1. Data analysis technique was appropriate	5	T-test was used to analyze the data.
7.2. Effect size was reported	5	The researchers reported the effect size of the study.
Total Scores	**139** out of 145	
Total Percent	**95.9%**	

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008).
Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.
N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

Quantitative Research- Lowest Scores Received

Author of Study	Ratnasari et al., 2017

Indicators	Scores	Justifications for Scores Given
1. Research Conceptualization		
1.1. Rationale of study was stated	5	The researchers emphasized the need for teaching creative thinking skills to students in a vocational school by incorporating it with PBL.
1.2. Purpose of study was stated	5	The researchers examined the effect of the use of PBL on students' creativity in grade 7 in a vocational school.
1.3. Theoretical framework was discussed	1	No theoretical framework was discussed.
1.4. Research questions were stated	1	The researchers did not state the research question.
2. Settings and Participants		
2.1. Description of the setting was provided	2	The researchers stated only that the study was conducted in a vocational high school in one grade 7 classroom.
2.2. Description of how the setting was selected was provided	1	No information was stated.
2.3. Procedures for selecting participants were discussed	1	The researchers did not mention how they selected the sample of this study.
2.4. Relevant demographic information about participants was provided	2	Only the number of students was stated (32 students participated in this study).
2.5. Description of the role of researcher(s) in relation to the settings and participants was clear	1	No information was stated.
3. Intervention		
3.1. Relevant demographic information of interventionists was provided	1	No information was stated.
3.2. Interventionists' roles were described	5	The researchers stated, "The teacher acts as a facilitator and assesses student performance through the observation sheet."
3.3. Interventionists were equivalent across conditions	N/A	In this study, one group was included. The researchers used a group posttest design in their study.

(Continued)

Author of Study	Ratnasari et al., 2017

Indicators	Scores	Justifications for Scores Given
3.4. Description of the practice was stated clearly	1	The researchers did not describe the practice and gave no overall picture about PBL.
3.5. Description of how the practice was implemented was included	3	Overall, the researchers reported that PBL was implemented to teach three concepts: dyes, preservatives/antioxidants, and sweetener. However, they did not discuss in detail how the practice was implemented.
3.6. Relevant aspects of the practice in the control group were described	N/A	No control group was included in this study.
4. Fidelity of Implementation (FOI)		
4.1. FOI was assessed	1	No information was stated.
4.2. Key features of the practice were assessed	1	No information was stated.
4.3. Inter-observer reliability of FOI was collected regularly	1	No information was stated.
5. Attrition		
5.1. Overall attrition was low across groups	N/A	Only one group was included in this study.
5.2. Difference in percentage of attrition between groups was low	N/A	Only one group was included in this study.
6. Measurements		
6.1. Multiple outcome measures were used	5	Three instruments were implemented: an essay, a student worksheet, and observations.
6.2. Data collection was described	1	No information was stated about how the researchers collected data.
6.3. At least one measure of generalized performance was used	1	None of the instruments were considered valid to generalize the results, due to the lack of information about validity and reliability.
6.4. Outcomes were measured at both pretest and posttest, at a minimum	1	A posttest only was design.
6.5. Outcome measures had evidence of internal consistency	1	No information was stated.

(Continued)

Author of Study	Ratnasari et al., 2017

Indicators	Scores	Justifications for Scores Given
6.6. Inter-observer reliability was assessed and acceptable	1	No information was stated.
6.7. Evidence of concurrent, content, construct, and/or predictive validity was provided	1	No information was stated.
7. Data Analysis		
7.1. Data analysis technique was appropriate	1	The researchers reported only percentages in their results. However, relying only on percentage is not sufficient to provide scientific evidence for the value of the intervention.
7.2. Effect size was reported	1	No information was stated.
Total Scores	**45** out of 125	
Total Percent	**36%**	

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008).

Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.

N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

Qualitative Research- Highest Scores Received

Author of Study	Wu et al,, 2015

Indicators	Scores	Justification for Scores Given
1. Research Conceptualization		
1.1. Rationale of study was stated	5	The researchers expressed the importance of engaging students in learning situations to increase their desire for learning.
1.2. Purpose of study was stated	5	The researchers investigated the participants' perceptions about the REAPS model.
1.3. Theoretical framework was discussed	5	The REAPS model was the theoretical framework for this study.
1.4. Research questions were stated	5	One research question guided this study.
2. Settings and Participants		
2.1. Description of the setting was provided	5	Sufficient information was provided about the school and services that were available for students.
2.2. Description of how the setting was selected was provided	5	The school was a place to implement a large research project and examine the effect of the use of REAPS.
2.3. Procedures for selecting participants were discussed	5	The participants were selected randomly for interviews across classrooms to participate in this study.
2.4. Relevant demographic information about participants was provided	5	Sufficient information was provided about the participants.
2.5. Description of the role of researcher(s) in relation to the settings and participants was clear	5	One of the researchers and two school staff conducted pretests of creative problem-solving in science to all students in the school.
3. Intervention		
3.1. Relevant demographic information of interventionists were provided	1	No Information was provided.

(Continued)

Author of Study	Wu et al,, 2015

Indicators	Scores	Justification for Scores Given
3.2. Interventionists' roles were described	3	The teachers participated in a professional development workshop to implement REAPS in their classrooms and to develop their teaching units.
3.3. Description of the practice was stated clearly	5	The researchers described the REAPS model in the introduction of their study.
3.4. Description of how the practice was implemented was included	3	The researchers described the practice in general without sufficient details.
4. Data Sources and Analysis		
4.1. Data collection was described	5	The researchers described data collection in detail.
4.2. Multiple data sources were included	5	Interviews and artifacts were the data for this study.
4.3. Framework for data analysis was described	5	The researchers described the framework for data analysis in detail.
4.4. Coding procedures were described	5	The process the researchers used for coding data was reported clearly.
4.5. Identification of themes was described	5	The process the researchers used for identifying themes was reported clearly.
5. Trustworthiness and Credibility		
5.1. Trustworthiness and credibility were established	5	The researchers asked an external person to participate in the data analysis process.
5.2. Audit trail was clear and detailed enough to follow	5	The researchers described the audit trail clearly.
6. Outcome(s)		
6.1. Description of outcomes was clear	5	The researchers reported the findings in depth.

(Continued)

Author of Study	Wu et al,, 2015

Indicators	Scores	Justification for Scores Given
6.2. Strength of the outcomes or effects was stated	5	The researchers highlighted the major findings and the strength of the results.
Total Scores	**102** Out of 110	
Total Percent	**92.7%**	

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008). Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.
N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

Qualitative Research- Lowest Scores Received

Author of Study	Nurdyani et al., 2018

Indicators	Scores	Justification for Scores Given
1. Research Conceptualization		
1.1. Rationale of study was stated	5	The researchers emphasized the need for improving students' creativity in mathematics.
1.2. Purpose of study was stated	5	The authors described the level of creative thinking in mathematics among gifted students in grade 8 after exposing them to PBL.
1.3. Theoretical framework was discussed	1	No information was discussed.
1.4. Research questions were stated	1	The researchers did not state the research question.
2. Settings and Participants		
2.1. Description of the setting was provided	1	No information was stated.
2.2. Description of how the setting was selected was provided	2	The researchers did not provide any information about the setting other than that the study was conducted with students in grade 8.
2.3. Procedures for selecting participants were discussed	1	The method of selecting students in this study was not appropriate; the researchers used problems about the coordinate system, and students who received a score of 80% were chosen as highly-able students in math.
2.4. Relevant demographic information about participants was provided	2	Only the number of students (36) was stated.
2.5. Description of the role of researcher(s) in relation to the settings and participants was clear	1	No information was stated.
3. Intervention		
3.1. Relevant demographic information of interventionists was provided	1	No information was stated.
3.2. Interventionists' roles were described	5	The teacher worked as a facilitator.

(Continued)

Author of Study	Nurdyani et al., 2018

Indicators	Scores	Justification for Scores Given
3.3. Description of the practice was stated clearly	5	The researchers described PBL in the introduction of their study.
3.4. Description of how the practice was implemented was included	3	The researchers provided an overall description of how PBL was implemented. They mentioned that students were taught relation and function concepts in mathematics.
4. Data Sources and Analysis		
4.1. Data collection was described	2	The researchers used three assessments: observation, a math test, and interviews. However, they did not describe the data collection in a way that can be replicated.
4.2. Multiple data sources were included	5	Three instruments were used: observation, tests, and interviews.
4.3. Framework for data analysis was described	2	Three assessments were used: observation, a math test, and interviews. However, there was no clear discussion of these instruments. For example, no validity and reliability information was reported.
4.4. Coding procedures were described	1	No information was stated about how the researcher analyzed their interviews and observation data, and how they coded the data.
4.5. Identification of themes was described	1	No information was stated.
5. Trustworthiness and Credibility		
5.1. Trustworthiness and credibility were established	1	No information was stated.
5.2. Audit trail was clear and detailed enough to follow	1	No information was stated.
6. Outcome(s)		
6.1. Description of outcomes was clear	3	The researchers highlighted the major findings. However, they should have described in detail how they interpreted the observation and interview data.

(Continued)

Author of Study	Nurdyani et al., 2018

Indicators	Scores	Justification for Scores Given
6.2. Strength of the outcomes or effects was stated	5	The researchers pointed out that students' creative thinking in mathematics as measured by fluency, flexibility, and originality was high after implementing PBL.
Total Scores	**54** Out of 110	
Total Percent	**49.1%**	

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008).

Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.

N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

Mixed-Methods Research- Highest Scores Received

Author(s) of Study	Siew et al., 2015

Indicators	**Scores**	**Justification for Scores Given**
1. Research Conceptualization		
1.1. Rationale of study was stated	5	The researchers emphasized the need for investigating the effect of the use of PBL on students' scientific creativity.
1.2. Purpose of study was stated	5	The researchers investigated whether the use of PBL increased scientific creativity among students in grade 5.
1.3. Theoretical framework was discussed	5	The theoretical framework was based on the Scientific Creativity Structure Model (Hu & Adey, 2002).
1.4. Research questions were stated	5	Three research questions guided this study.
2. Settings and Participants		
2.1. Description of the setting was provided	5	The researchers stated that the study was conducted in two elementary schools in seven classes. Each class was comprised of 30 students. The schools had moderate science achievement. The researchers also stated the occupations of students' parents.
2.2. Description of how the setting was selected was provided	5	The researchers received funding to implement the study in these schools.
2.3. Procedures for selecting participants were discussed	5	Purposive sampling was used to select the participants.
2.4. Relevant demographic information about participants was provided	5	The number of students, their gender, and their families' occupations were reported.
2.5. Description of the role of researcher(s) in relation to the setting and participants was clear	5	The researchers received funding to implement the study in these schools.

(Continued)

Author(s) of Study	Siew et al., 2015

Indicators	Scores	Justification for Scores Given
3. Intervention		
3.1. Interventionists' roles were described	5	The teachers were facilitators in instructional processes.
3.2. Relevant demographic information of interventionists was provided	1	No information was provided about the teachers.
3.3. Interventionists were equivalent across conditions	N/A	No control group was included in this study.
3.4. Description of the practice was stated clearly	5	The researchers described PBL in the introduction of their study.
3.5. Description of how the practice was implemented was included	5	The researchers provided a description of how the intervention was implemented in classrooms. The intervention included five lessons plan.
4. Fidelity of Implementation (FOI)		
4.1. FOI was assessed	1	No information was provided about assessing FOI during the intervention.
4.2. Key features of the practice were assessed	1	No information was provided.
4.3. Inter-observer reliability of FOI was collected regularly	1	No information was provided.
5. Measurements		
5.1. Data collection was described	5	Pretests and posttests were used to collect the data before and after the intervention. Also, students completed open-ended questions at the end of the intervention.
5.2. Multiple data sources were included	5	Two instruments were used: open-ended questions and a scientific creativity test.
5.3. Outcome measures had evidence of internal consistency	5	The internal consistency was reported.
5.4. Inter-observer reliability was assessed and acceptable	N/A	

(Continued)

Author(s) of Study	Siew et al., 2015

Indicators	Scores	Justification for Scores Given
5.5. Evidence of concurrent, content, construct, and/or predictive validity was provided	3	The face validity was reported. The correlation coefficient between Form A and B in the test was reported. However, the correlation was .30 between the two forms. The researchers did not report the construct and/or predictive validity of the scientific test.
6. Data Analysis		
6.1. Data analysis technique was appropriate	5	A t-test was used for analyzing quantitative data. Coding and categorizing were used for qualitative data.
6.2. Coding procedures were described	5	The researcher coded students' responses in open-ended questions.
6.3. Identification of themes was described	5	The researchers used interpretive methods (Erickson, 1986) to arrive at themes.
7. Trustworthiness and Credibility		
7.1. Trustworthiness and credibility were established	5	A science teacher and one researcher scored students' answers independently. Both of the raters attended a professional development workshop to introduce the components of the test.
7.2. Audit trail was clear and detailed enough to follow	5	The procedures for scoring were clear.
8. Outcome(s)		
8.1. Description of outcomes was clear	5	The researchers discussed the outcomes in detail.
8.2. Strength of the outcomes or effects was stated.	5	The researchers highlighted the major findings in their study.
8.3. At least one measure of generalized performance was used	5	The procedure for scoring students' answers was appropriate, so the results can be generalized.

(Continued)

Author(s) of Study	Siew et al., 2015

Indicators	Scores	Justification for Scores Given
8.4. Effect size was reported	1	The effect size was not reported.
Total Scores	**123** out of 145	
Total Percent	**84.8%**	

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008). In the CEC, only quality indicators for qualitative and quantitative research were developed. I combined the quality indicators of both qualitative and quantitative research for the mixed-methods research design.
Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.
N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

Mixed-Methods Research - Lowest Scores Received

Author(s) of Study	West, 2008

Indicators	Scores	Justification for Scores Given
1. Research Conceptualization		
1.1. Rationale of study was stated	1	The researcher did not justify the importance of her study.
1.2. Purpose of study was stated	5	The researcher investigated the use of TASC and its effectiveness on students' thinking skills and motivation of learning.
1.3. Theoretical framework was discussed	5	The theoretical framework was based on the TASC model.
1.4. Research questions were stated	1	No research questions were stated. The researcher stated only the purpose of the study.
2. Settings and Participants		
2.1. Description of the setting was provided	2	No descriptive information was reported. The only information stated was the number of schools (three).
2.2. Description of how the setting was selected was provided	1	No information was reported.
2.3. Procedures for selecting participants were discussed	1	No information was reported.
2.4. Relevant demographic information about participants was provided	2	The researcher did not report the number of students. She stated their ages and that students were of mixed abilities and mixed genders.
2.5. Description of the role of researcher(s) in relation to the setting and participants was clear	1	No information was reported.
3. Intervention		
3.1. Interventionists' roles were described	1	No information was reported.
3.2. Relevant demographic information of interventionists was provided	1	No information was reported.
3.3. Interventionists were equivalent across conditions	N/A	
3.4. Description of the practice was stated clearly	1	No information was reported.

(Continued)

Author(s) of Study	West, 2008

Indicators	Scores	Justification for Scores Given
3.5. Description of how the practice was implemented was included	1	No information was reported.
4. Fidelity of Implementation (FOI)		
4.1. FOI was assessed	1	No information was reported.
4.2. Key features of the practice were assessed	1	No information was reported.
4.3. Inter-observer reliability of FOI was collected regularly	1	No information was reported.
5. Measurements		
5.1. Data collection was described	3	The researcher used four instruments in her study, which may be considered sufficient. However, the researcher did not describe these instruments.
5.2. Multiple data sources were included	5	Four instruments were administered.
5.3. Outcome measures had evidence of internal consistency	N/A	
5.4. Inter-observer reliability was assessed and acceptable	N/A	
5.5. Evidence of concurrent, content, construct, and/or predictive validity was provided	N/A	
6. Data Analysis		
6.1. Data analysis technique was appropriate	1	No information was reported.
6.2. Coding procedures were described	1	No information was reported.
6.3. Identification of themes was described	1	No information was reported.
7. Trustworthiness and Credibility		
7.1. Trustworthiness and credibility were established	1	No information was reported.
7.2. Audit trail was clear and detailed enough to follow	1	No information was reported.

(Continued)

Author(s) of Study	West, 2008

Indicators	Scores	Justification for Scores Given
8. Outcome(s)		
8.1. Description of outcomes was clear	3	The researcher discussed her findings. However, she did not report her method for analyzing data.
8.2. Strength of the outcomes or effects was stated	3	The researcher indicated the importance of her findings but included no description of the way she arrived at her conclusions.
8.3. At least one measure of generalized performance was used	1	None of the instruments can be used to generalize the findings of the study.
8.4. Effect size was reported	N/A	
Total Scores	**46** out of 130	
Total Percent	**35.4%**	

Note: These indicators were modified from the Council for Exceptional Children, Practice Studies Manual (2008). In the CEC, only quality indicators for qualitative and quantitative research were developed. I combined the quality indicators of both qualitative and quantitative research for the mixed-methods research design.
Scores given ranged from 1 to 5. 1 = Inadequate/Unknown, 2 = weakly stated, 3 = Partial, 4 = good, and 5 = Adequate.
N/A = Not Applicable. Sub-indicators that were given N/A were excluded from the calculation of total scores and percents.

APPENDIX G

CONSENT FORM FOR PARENTS IN GROUP A CLASSES

CONSENT FORM FOR PARENTS IN GROUP A CLASSES

Research Project: Real Engagement in Active Problem Solving: Differentiation for Diverse Learners in Regular Classrooms

I (*print name*)______________________________________give consent to the participation of

my child (*print name*) __________________________ in the research project described below.

TITLE OF THE PROJECT: Real Engagement in Active Problem Solving: Differentiation for Diverse Learners in Regular Classrooms

CHIEF RESEARCHER: [redacted] or [redacted]

CO-RESEARCHERS: [redacted]

xxxx[redacted]

In giving my consent I acknowledge that

1. The procedures required for the project and the time involved have been explained to me and any questions I have about the project have been answered to my satisfaction.

2. I have read the Parent Information Sheet and have been given the opportunity to discuss the information and my child's involvement in the project with the researchers.

3. I have discussed participation in the project with my child and my child assents to their participation in the project.

4. I understand that that my child's participation in this project is voluntary; a decision not to participate will in no way affect her or his academic standing or relationship with the school and my child is free to withdraw his or her participation in the research (analysis of test results) at any time.

5. I understand that my child's involvement is strictly confidential and that no information about my child will be used in any way that reveals my child's identity.

6. I understand that my child's test results will be analyzed to determine the effects of a special program, and that these test results will be stored in a place and in a way that protects my child's anonymity.

7. I understand that audio recordings will be made as part of the study if my child is selected for an interview. These recordings will take place during the school day and in a quiet area of the classroom. Other children will be in the classroom, but not interfering with my child's interview.

Please cross out any activity that you do not wish your child to participate in.

Signed……………………………………………………………………………..

Name……………………………………………………………………………….

Date………………………………………………………………………………….

Any complaints about the research may be registered with the Director Educational Services, [redacted]

APPENDIX H

PARENT/CAREGIVER INFORMATION SHEET

PARENT/CAREGIVER INFORMATION SHEET

Research Project: Real Engagement in Active Problem Solving: Differentiation for Diverse Learners in Regular Classrooms

You and your child are invited to take part in a study being conducted by [redacted]
It is part of a project designed to adapt, modify, and test the effectiveness of an educational teaching program developed in the USA.

We are asking you if it is okay for you and your child to take part in this project.

We are trying to find out how the teaching program needs to be changed (if at all) to make it more useful and appropriate in Australia, and to find out if the use of this teaching method can increase students' creativity (general, scientific, and mathematical), their understanding of science concepts, and their achievement. We want to know if the use of this approach helps teachers differentiate their instruction to meet the needs of diverse learners, especially those who are gifted.

The information from the study will be used to decide whether to use this teaching method in [redacted] how to adapt or change it to fit our school community, and whether to teach others how to implement it. We will report the results to you at parent meetings and will distribute a report to you. We also will report the results in educational journals, at educational conferences, and in teacher workshops.

We will ask your child to take some educational tests that are very similar to tests they usually take in school: a test of general creativity in which they will complete an incomplete picture; a test of scientific creativity in which they will identify science problems and design ways to solve them; a test of mathematical creativity in which they will solve several math problems and create math problems; and a concept mapping exercise in which they will draw diagrams of the ways science ideas are related. The tests will take approximately 2 hours (over a period of one week) and will occur twice each year. Your child also will take the usual tests of learning progress given at certain times during the school year.

Because your child is in a class with a number of students who are on Individualised Learning Plans, your child also will experience the new educational method, which will include solving real life problems individually and in groups. These will be part of the normal teaching of all subjects, but may be used mostly in science classes.

Participation in the classes and the educational program is a normal part of your child's education, but your child's participation in the research is voluntary. Your child will take the tests as part of the normal gathering of information about children's abilities, but the researchers will analyze the results of your child's tests only if both you and your child agree. If you decide not to take part, it will not affect your child's results or progress at school, and if you or your child change your mind about taking part, even after the study has started, just let the teacher and the school principal know, and any information already collected about your child will not be used in the research. Please notify the principal, [redacted], and your child's teacher if you wish to withdraw your child from the project.

If you and your child decide to participate, no one will be able to identify you or your child from the results of the study. Only the researchers will have access to this information, and it will not have identifying information about you or your child. The test scores will be stored on a password-protected computer and all test materials will be stored in a protected file in the primary researcher's office. These scores and materials will be kept for as long as the project lasts.

Six children from each classroom will be interviewed about their experiences as part of this program. They will make drawings and will talk about their drawings. Audio recordings of your child's responses will be made. As soon as they are transcribed, the audio recordings will be deleted. The transcriptions will not have your child's name on them. Only the interviewer and the individual who transcribes the recordings will hear your child's voice.

These recordings will be collected on a date to be specified, and the researchers will interview your child individually in a quiet area of the classroom while other children are present. Only the researchers will have access to the transcriptions of the recordings and the child's drawings.

We will use the recordings to find out how children think and feel about their participation in the new educational program. We believe that the children's perceptions are important in making decisions about the value of new teaching methods.

If your child is selected for an interview, we also will want to interview you about your perceptions of your child's participation in the program. Audio recordings of your responses will be made. As soon as they are transcribed, the audio recordings will be deleted. The transcriptions will not have your name on them. Only the interviewer and the individual who transcribes the recordings will hear your voice.

These recordings will be collected on a date to be specified, and the researchers will interview you individually in a place that is quiet and comfortable for you. Only the researchers will have access to the transcriptions of the recordings.

We will use the recordings to find out how parents think and feel about their child's participation in the new educational program. We believe that children's and parent's perceptions are important in making decisions about the value of new teaching methods.

You also will be asked to respond to questionnaires about the use of the new teaching method. When questionnaires are distributed, by responding to the questionnaire, you will be giving your permission for your responses to be analyzed by the researchers.

If you would like to check that you are OK with the information or recordings from the study, or if you have any questions, please ask the principal_[redacted], about the process.

When you have read this information, [redacted] will be available to answer any questions you may have at this time. If you would like to know more at any stage, please feel free to contact:

[redacted]

[redacted]

This information sheet is for you to keep.

APPENDIX I

THE FORM FOR OBSERVING IMPLEMENTATION BY THE TEACHERS IN CLASSROOMS

Real Engagement in Active Problem Solving
Fidelity of Implementation

Observer: __________________________ Today's Date: _______________________
Teacher: __________________________ Grade/Year: _________________________
School: __________________________ Section/Class: ________________________
Number of Students: ____
Beginning Time: ______________________ Ending Time: _______________________
Teacher's Unit/Lesson Plan Attached: ____

<u>Prism Abilities</u>
__Logical/Mathematical __Somatic/Bodily __Visual/Spatial __Auditory __Linguistic
__Mechanical/Technical __Scientific __Social __Emotional __Spiritual

DISCOVER Problem Types
______Type I _____Type II ______Type III ______Type IV ______Type V ______Type VI

REAPS Project Title:

REAPS Problem Synopsis and Problem Type(s):

REAPS TASC Step(s): Gather/Organize, Identify, Generate, Decide, Implement, Evaluate, Communicate, Learn from Experience

PBL Stakeholder Groups:

Observation Notes

<u>Content</u>
__Abstractness
__Complexity
__Variety
__Organization
__Study of People
__Study of Methods

Observer: ________________________ Today's Date: ________________________
Teacher: ___________________________ Grade/Year: _________________________

<u>Process</u>
__Higher Levels of Thinking
__Open-Endedness
__Discovery
__Evidence of Reasoning
__Freedom of Choice
__Pacing
__Variety
__Group Interaction

<u>Product</u>
__Real Problems
__Real Audience
__Transformation
__Variety
__Self-Selected Formats
__Appropriate Evaluation
__Photos of Product Attached

Observer: ____________________________ Today's Date: _________________________
Teacher: ____________________________ Grade/Year: __________________________

Learning Environment	**Description** (Attach photographs of the classroom)
__Learner Centered	
__Encourage Independence	
__Openness	
__Accepting	
__Complexity	
__Various Groupings	
__Flexibility	
__High Mobility	

APPENDIX J

SELF-EVALUATION FORM

Checklist for Assessment of Implementation of the Real Engagement in Active Problem Solving (REAPS) Teaching Model

C. June Maker, Robert Zimmerman, Randy Pease, and Abdulnasser Alhusaini

Teacher ______________________ Evaluator ____________________ Date ____________

School ______________________ Year (Grade level) __________

Please give yourself a rating from 0 to 6 on each behavior on the checklist. In the areas with several alternatives, please check the ones that you do. If you would like to explain a rating, please do so in the boxes beside the behavior or on the back of the page. Please use the following scale to make your ratings:

0: I don't know how to do this.

1: I know how to do this, but I never do it.

2: I seldom do this.

3: I sometimes do this.

4: I often do this.

5: I always do this when I do REAPS.

6: This is an integral part of my teaching practice whether or not I am doing REAPS.

1. Content		
Principles and Comments	**Model**	**Rating**
1.1. Abstractness		
1.1.1. A problem is posed that has a Macro Concept as the overall focus for problem solving.	PBL	
1.1.2. The key question for the teaching unit involving the problem is based on the Macro Concept(s).	PBL	
1.1.3. Content is organized around broad-based interdisciplinary themes, called Macro Concepts in the framework.	DISCOVER	
1.6. Study of Methods		
1.6.1 Students collect data using the methods appropriate for the type of problem being solved.	PBL	
1.6.2. Students employ research or observational techniques appropriate to the discipline.	PBL	

1.6.3. Problems or tasks are completed using creative and productive thinking in a systematic way.	TASC	
1.6.4. The teacher uses the problem solving model across disciplines and projects.	TASC	
1.6.5. The teacher demonstrates a variety of processes (across abilities and content areas).	DISCOVER	
1.6.6. The teacher gives students opportunities to use a variety of processes.	DISCOVER	
1.6.7. The teacher uses the problem continuum: • two types (I and II) in which students learn to use specific methods ____ • four types (III, IV, V, & VI) in which they select, create, and apply their own methods ____	DISCOVER	
2. Processes		
2.1. Higher Level Thinking		
2.1.1. Students gather information using appropriate methods.	PBL	
2.1.2. Students apply the information.	PBL	
2.1.3. Students develop new and different solutions.	PBL	
2.1.4. The *gather and organize* step includes methods such as • interviewing experts and community members ____ • naturalistic observations ____ • designing, conducting, and analyzing results of experiments ____	TASC	
2.1.5. A sufficient amount of time is spent on each step of the problem solving process: • *gather and organize* ____ • *identify* ____ • *generate* ____ • *decide* ____ • *implement* ____ • *evaluate* ____ • *communicate* ____ • *reflect* ____	TASC	
2.1.6. Higher levels of thinking are emphasized in all problem types: • I ____ • II ____ • III ____ • IV ____ • V ____ • VI ____	DISCOVER	
2.2. Open-Endedness		
2.2.1. The problems chosen are provocative, persistent, or challenging.	PBL	
2.2.2. Various stakeholders or those with divergent perspectives may propose different solutions.	PBL	
2.2.3. When they *Gather and Organize*, students list what they know, what they want to know, and where they might find this information.	TASC	

2.2.4. At the *Identify* step, students focus on defining the task or problem.	TASC	
2.2.5. At the *Generate* step, students generate as many ideas as possible to solve the problem.	TASC	
2.2.6. When students *Decide*, they do so on the basis of criteria they have developed.	TASC	
2.2.7. Students *Implement* the solution they have devised.	TASC	
2.2.8. Students *Evaluate* the solution based on the criteria they have developed.	TASC	
2.2.9. Students *Communicate* their solutions using their own presentation skills or products.	TASC	
2.2.10. Students *Reflect* on what they learned from their experience.	TASC	
2.2.11. A sufficient amount of time is spent on solving open-ended problems: • Type V ____ • Type VI ____	DISCOVER	
3. Products		
3.1. Result from Real Problems		
3.1.1. Students are given problems to resolve that are real • in their own lives ____ • in their communities and regions ____ • in the world ____	PBL	
3.2. Addressed to Real Audiences		
3.2.1. Results of problem solving experiences are communicated to an audience interested in the solution and • delivered in the form of a report ____ • presented as a talk ____ • given as a PowerPoint presentation ____ • presented as a movie ____ • presented in another way ____	PBL	
3.2.2. Students present to • peers ____ • parents ____ • students in other classrooms ____	PBL	
3.2.3. The teacher arranges ways for students to communicate to audiences interested in their products.	TASC	
4. Learning Environment		
4.1. Learner-Centered		
4.1.1. The teacher is the facilitator while students solve the problem.	PBL	
4.1.2. The teacher determines the problem situation, but students solve the problem themselves.	PBL	
4.1.3. When possible, the teacher asks students to choose local, real problems they are interested in solving.	PBL	

4.1.4. As students follow the steps in problem solving, the teacher guides rather than directs the students to • *gather and organize* ____ • *identify* ____ • *generate* ____ • *decide* ____ • *implement* ____ • *evaluate* ____ • *communicate* ____ • *reflect* ____	TASC	
4.1.5. The teacher establishes a learner-centered environment that includes • flexible scheduling ____ • flexible grouping ____ • standards for behavior ____ • sharing ____ • openness____ • acceptance ____	DISCOVER	
4.2. Encourages Independence		
4.2.1. Students are encouraged to solve the problems in their own ways.	PBL	
4.2.2. Students are encouraged to identify problems that are important to them.	PBL	
4.2.3. Students are taught the problem solving process so they can use it on their own.	TASC	
4.2.4. Students are encouraged to apply the problem solving process to problems or tasks in academic or social contexts.	TASC	
4.2.5. Students are encouraged to • identify their own problems ____ • select the formats of their products ____ • produce varied products that reflect their individuality ____	DISCOVER	
5. General Comments		
General Comments (cont'd)		

APPENDIX K

CHECKLIST FOR ASSESSMENT OF IMPLEMENTATION OF THE REAL ENGAGEMENT IN ACTIVE PROBLEM SOLVING (REAPS) TEACHING MODEL BY SUPERVISORS

Checklist for Assessment of Implementation of the Real Engagement in Active Problem Solving (REAPS) Teaching Model

C. June Maker, Robert Zimmerman, Randy Pease, and Abdulnasser Alhusaini

Teacher ____________________ Evaluator __________________ Date ___________

School ____________________ Year (Grade level) _________

Please give a rating from 0 to 6, with 0 = no evidence, 1 = low, 2 = below average, 3 = average, 4 = above average, 5 = high, and 6 = excellent, on each behavior. Make a check in the blanks, when provided, to indicate which of the alternatives apply. When appropriate, provide comments to explain your ratings.

1. Content		
Principles and Comments	**Model**	**Rating**
1.1. Abstractness		
1.1.1. A problem is posed that has a Macro Concept as the overall focus for problem solving.	PBL	
1.1.2. The key question for the teaching unit involving the problem is based on the Macro Concept(s).	PBL	
1.1.3. Content is organized around broad-based interdisciplinary themes, called Macro Concepts in the framework.	DISCOVER	
1.6. Study of Methods		
1.6.1 Students collect data using the methods appropriate for the type of problem being solved.	PBL	
1.6.2. Students employ research or observational techniques appropriate to the discipline.	PBL	
1.6.3. Problems or tasks are completed using creative and productive thinking in a systematic way.	TASC	
1.6.4. The teacher uses the problem solving model across disciplines and projects.	TASC	
1.6.5. The teacher demonstrates a variety of processes (across abilities and content areas).	DISCOVER	
1.6.6. The teacher gives students opportunities to use a variety of processes.	DISCOVER	
1.6.7. The teacher uses the problem continuum:	DISCOVER	

• two types (I and II) in which students learn to use specific methods ____ • four types (III, IV, V, & VI) in which they select, create, and apply their own methods ____		
2. Processes		
2.1. Higher Level Thinking		
2.1.1. Students gather information using appropriate methods.	PBL	
2.1.2. Students apply the information.	PBL	
2.1.3. Students develop new and different solutions.	PBL	
2.1.4. The gather and organize step includes methods such as • interviewing experts and community members ____ • naturalistic observations ____ • designing, conducting, and analyzing results of experiments ____	TASC	
2.1.5. A sufficient amount of time is spent on each step of the problem solving process: • gather and organize ____ • identify ____ • generate ____ • decide ____ • implement ____ • evaluate ____ • communicate ____ • reflect ____	TASC	
2.1.6. Higher levels of thinking are emphasized in all problem types: • I ____ • II ____ • III ____ • IV ____ • V ____ • VI ____	DISCOVER	
2.2. Open-Endedness		
2.2.1. The problems chosen are provocative, persistent, or challenging.	PBL	

2.2.2. Various stakeholders or those with divergent perspectives may propose different solutions.	PBL	
2.2.3. When they Gather and Organize, students list what they know, what they want to know, and where they might find this information.	TASC	
2.2.4. At the Identify step, students focus on defining the task or problem.	TASC	
2.2.5. At the Generate step, students generate as many ideas as possible to solve the problem.	TASC	
2.2.6. When students Decide, they do so on the basis of criteria they have developed.	TASC	
2.2.7. Students Implement the solution they have devised.	TASC	
2.2.8. Students Evaluate the solution based on the criteria they have developed.	TASC	
2.2.9. Students Communicate their solutions using their own presentation skills or products.	TASC	
2.2.10. Students Reflect on what they learned from their experience.	TASC	
2.2.11. A sufficient amount of time is spent on solving open-ended problems: • Type V ____ • Type VI ____	DISCOVER	
3. Products		
3.1. Result from Real Problems		
3.1.1. Students are given problems to resolve that are real • in their own lives ____ • in their communities and regions ____ • in the world ____	PBL	
3.2. Addressed to Real Audiences		
3.2.1. Results of problem solving experiences are communicated to an audience interested in the solution and • delivered in the form of a report ____ • presented as a talk ____ • given as a PowerPoint presentation ____ • presented as a movie ____	PBL	

• presented in another way ____		
3.2.2. Students present to • peers ____ • parents ____ • students in other classrooms ____	PBL	
3.2.3. The teacher arranges ways for students to communicate to audiences interested in their products.	TASC	
4. Learning Environment		
4.1. Learner-Centered		
4.1.1. The teacher is the facilitator while students solve the problem.	PBL	
4.1.2. The teacher determines the problem situation, but students solve the problem themselves.	PBL	
4.1.3. When possible, the teacher asks students to choose local, real problems they are interested in solving.	PBL	
4.1.4. As students follow the steps in problem solving, the teacher guides rather than directs the students to • gather and organize ____ • identify ____ • generate ____ • decide ____ • implement ____ • evaluate ____ • communicate ____ • reflect ____	TASC	
4.1.5. The teacher establishes a learner-centered environment that includes • flexible scheduling ____ • flexible grouping ____ • standards for behavior ____ • sharing ____ • openness____ • acceptance ____	DISCOVER	
4.2. Encourages Independence		

4.2.1. Students are encouraged to solve the problems in their own ways.	PBL	
4.2.2. Students are encouraged to identify problems that are important to them.	PBL	
4.2.3. Students are taught the problem solving process so they can use it on their own.	TASC	
4.2.4. Students are encouraged to apply the problem solving process to problems or tasks in academic or social contexts.	TASC	
4.2.5. Students are encouraged to • identify their own problems ____ • select the formats of their products ____ • produce varied products that reflect their individuality ____	DISCOVER	
4.5.2. Students are given the freedom to access materials and tools • within the classroom ____ • outside the classroom ____	DISCOVER	

5. General Comments

REFERENCES

Abbott, M. L. (2017). *Using statistics in the social and health sciences with spss® and excel* (1st ed.). Hoboken, New Jersey: John Wiley & Sons.

Abu Awwad, F. M., Asha, I. K., & Jado, S. M. (20pp14). The effect of TASC wheel on developing self-directed learning readiness and academic self-efficacy on a sample of 7th graders in Jordan. *Education, 135*(2), 237-251.

Acock, A. (2014). A gentle introduction to Stata (4th ed.).College Station, Texas: Stata Press.

Adams, H. B., & Wallace, B. (1991). TASC: A model for curriculum which could have application in a wide variety of social, economic and political situations. Developing the potential of children in disadvantaged communities: The TASC project: "thinking actively in a social context". *Gifted Education International*, *7*(3), 104-113. doi: http://gei.sagepub.com/

Aidoo, B., Boateng, S. K., Kissi, P. S., & Ofori, I. (2016). Effect of problem-based learning on students' achievement in chemistry. *Journal of Education and Practice*, *7*(33), 103-108.

Akınoğlu, R. O., & Tandoğan, A. (2007). The effects of problem-based active learning in science education on students' academic achievement, attitude and concept learning. *Online Submission*, *3*(1), 71-81. doi:10.13140/RG.2.2.22857.03685

Alhusaini, A. (2016). *The effects of duration of exposure to the REAPS model in developing students' general creativity and creative problem solving in science.* (Doctoral Dissertation). Available from ProQuest Dissertation and Theses database. (Order No. 10113271). University of Arizona.

Alhusaini, A., Maker, C. J., & Alamiri, F. (2015). Adapting the REAPS model to develop students' creativity in Saudi Arabia: An exploratory study. Manuscript in preparation

Amabile, T. M. (1983). *The social psychology of creativity.* New York, NY: Springer-Verlag New York Inc.

Amabile, T. M. (1996). *Creativity in context: Update to the social psychology of creativity.* Boulder, CO: Westview Press Inc.

Amabile, T. M. (2013). Componential theory of creativity. In *Encyclopedia of management theory.* (pp. 134-139). Los Angeles, CA: SAGE Publications, Inc.

Anazifa, R. D., & Djukri, D. (2017). Project-based learning and problem-based learning: Are they effective to improve student's thinking skills?. *Jurnal Pendidikan IPA Indonesia*, *6*(2), 346-355. doi: 10.15294/jpii.v6i2.11100

Araz, G., & Sungur, S. (2007). Effectiveness of problem-based learning on academic performance in genetics. *Biochemistry and Molecular Biology Education, 35*(6), 448-451. doi:10.1002/bmb.97

Azano, A., Missett, T. C., Callahan, C. M., Oh, S., Brunner, M., Foster, L. H., & Moon, T. R. (2011). Exploring the relationship between fidelity of implementation and academic achievement in a third-grade gifted curriculum: A mixed-methods study. *Journal of Advanced Academics, 22*(5), 693-719. doi:10.1177/1932202X11424878

Bachelor, P., & Michael, W. B. (1991). Higher-order factors of creativity within Guilford's structure-of-intellect model: A re-analysis of a fifty-three variable data base. *Creativity Research Journal, 4*(2), 157-175. doi:10.1080/10400419109534383

Baer, J. (1993). *Creativity and divergent thinking: A task specific approach*. Hillsdale, N.J: L. Erlbaum.

Ball, S., & Henderson, K. (2008). Using the TASC wheel to challenge more able children in an inclusive environment. *Gifted Education International*, *24*(2-3), 258-261. doi: http://gei.sagepub.com/

Becker, H., & Geer, B. (1957). Participant observation and interviewing: A comparison. *Human organization*, 16(3), 28-32. doi:10.17730/humo.16.3.k687822132323013

Belland, B. R., Glazewski, K. D., & Richardson, J. C. (2011). Problem-based learning and argumentation: Testing a scaffolding framework to support middle school students' creation of evidence-based arguments. *Instructional Science, 39*(5), 667-694. doi:10.1007/s11251-010-9148-z

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., & Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology*, *23*(5), 443–451. doi:10.1037/0278-6133.23.5.443

Blandford, A. M., & Osher, F. C. (2012). *A checklist for implementing evidence-based practices and programs (EBPs) for justice-involved adults with behavioral health* *disorders*. Delmar, NY: SAMSHA's GAINS Center for Behavioral Health and Justice Transformation.

Burkus, D. (2014). *The myths of creativity: The truth about how innovative companies and people generate great ideas*. San Francisco, CA: Jossey-Bass.

Cartwright, S. (2008). A pilot project of 8 lessons across social studies: Thinking actively in a social context (TASC). *Gifted Education International*,*24*(2-3), 315-322. doi: http://gei.sagepub.com/

Council for Exceptional Children (2008). Classifying the state of evidence for special education professional practices: CEC practice study manual. Retrieved September 2, 2014, from https://www.cec.sped.org/

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd Ed). Thousand Oaks, CA: Sage publications.

Cropley, A. (2006). In praise of convergent thinking. *Creativity Research Journal, 18*(3), 391-404. doi:10.1207/s15326934crj1803_13

Csikszentmihalyi, M. (2006) A systems perspective on creativity. In: Henry J (3rd ed) *Creative management and development*. Sage Publications, London: UK, pp 3–17.

Csikszentmihalyi, M. (2014). *The systems model of creativity: The collected works of mihaly csikszentmihalyi*. Dordrecht: Springer Netherlands. doi: 10.1007/978-94-017-9085-7

Dane, A. V. (1997). *Program integrity in primary and early secondary prevention, preserving what works across diverse settings*. (Doctoral Dissertation). Available from National Library of Canada= Bibliothèque nationale du Canada.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, 18(1), 23-45. doi:10.1016/S0272-7358(97)00043-3

Davies, H. (2008). An overview of an investigation into the effects of using TASC strategies in the development of children's thinking and problem-solving skills in science. *Gifted Education International, 24*(2-3), 305-314. doi:10.1177/026142940802400323

Derzon, J. H., Sale, E., Springer, J. F., & Brounstein, P. (2005). Estimating intervention effectiveness: Synthetic projection of field evaluation results. *The Journal of Primary Prevention*, 26, 321–343. doi:10.1007/s10935-005-5391-5

Dods, R. (1997). An action research study of the effectiveness of problem-based learning in promoting the acquisition and retention of knowledge. *Journal for the Education of the Gifted, 20*(4), 423-437. doi:10.1177/016235329702000406

Drake, K. N., & Long, D. (2009). Rebecca's in the dark: A comparative study of problem-based learning and direct instruction/experiential learning in two 4th-grade classrooms. *Journal of Elementary Science Education, 21*(1), 1-16. doi:10.1007/BF03174712

DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American Journal of Community Psychology, 30*(2), 157-197. doi:10.1023/A:1014628810714

Duerden, M. D., & Witt, P. A. (2012). Assessing program implementation: What it is, why it's important, and how to do it. *Journal of Extension*, *50*(1), 1-8.

Dumas, J. E., Lynch, A. M., Laughlin, J. E., Smith, E. P., & Prinz, R. J. (2001). Promoting intervention fidelity: Conceptual issues, methods, and preliminary results from the early alliance prevention trial. *American Journal of Preventive Medicine, 20*(1), 38-47. doi: 10.1016/S0749-3797(00)00272-5

Durlak, J. A. (1998). Why program implementation is important. *Journal of Prevention & Intervention in the community*, *17*(2), 5-18. doi:10.1300/J005v17n02_02

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American journal of community psychology*, *41*(3-4), 327-350. doi:10.1007/s10464-008-9165-0

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health education research*, *18*(2), 237-256. doi: http://her.oxfordjournals.org/

Ersoy, E., & Baser, N. (2010). The effect of problem based learning process on student motivation. *Turkish Studies, 5*(4), 336-358. doi: http://www.turkishstudies.net/

Fan, X., Miller, B. C., Park, K., Winward, B. W., Christensen, M., Grotevant, H. D., & Tai, R. H. (2006). An exploratory study about inaccuracy and invalidity in adolescent self-report surveys. *Field Methods, 18(3)*, 223-244. doi:10.1177/152822X06289161

Fatimah, S. (2015). Devoting to enhance the critical thinking skill and the creativity of students in seventh grade through PBL model with JAS approachment. *Jurnal Pendidikan IPA Indonesia*, *4*(2), 149-157.

Faulkner, C. (2008). Creativity and thinking skills in mathematics: Using the TASC wheel as the basis for talented pupils to create their own thinking frameworks. *Gifted Education International, 24*(2-3), 288-296. doi: 10.1177/026142940802400321

Field, A. (2009). Discovering statistics using SPSS : And sex and drugs and rock 'n' roll. (3rd ed.). Los Angeles: SAGE.

Fitton, S., & Gilderdale, C. (2008). A case study of the use of TASC at Ollerton community primary school, UK. *Gifted Education International*,*24*(2-3), 217-224. doi: http://gei.sagepub.com/

Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on social work practice*, *19*(5), 531-540. doi:10.1177/1049731509335549

Fixsen, D. L., Naoom, S. F., Blase, K. A., & Friedman, R. M. (2005). *Implementation research: a synthesis of the literature*. Tampa, FL: Louis de la Parte Florida Mental Health Institute Publication.

Frey, B. B. (2016). *There's a stat for that: What to do & when to do it.* Thousand Oaks, CA: Sage Publications.

Fullan, M. (2001). *The new meaning of educational change* (3rd ed.). New York: Teachers College Press.

Gallagher, S. A., & Gallagher, J. J. (2013). Using problem-based learning to explore unseen academic potential. *Interdisciplinary Journal of Problem-based Learning,7*(1), 111-131. doi: http://dx.doi.org/10.7771/1541-5015.1322

Gallagher, S. A., Stepien, W. J., & Rosenthal, H. (1992). The effects of problem-based learning on problem solving. *Gifted Child Quarterly*, *36*(4), 195-200. doi: 10.1177/001698629203600405

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

George, D., & Mallery, P. (2016). *IBM SPSS Statistics 23 Step by Step : A Simple Guide and Reference*. London: Routledge.

Getzels, J., & Csikszentmihalyi, M. (1967). Scientific creativity. *Science Journal*, *3* (9), 80-84.

Getzels, J., & Csikszentmihalyi, M. (1976). *The creative vision: A longitudinal study of problem finding in art*. New York: Wiley.

Gibson, C., Folley, B. S., & Park, S. (2009). Enhanced divergent thinking and creativity in musicians: A behavioral and near-infrared spectroscopy study. *Brain and Cognition, 69*(1), 162-169. 10.1016/j.bandc.2008.07.009

Gomez-Arizaga, M. P., Bahar, A. K., Maker, C. J., Zimmerman, R., & Pease, R. (2016). How does science learning occur in the classroom? Students' perceptions of science instruction during the implementation of the REAPS model. *Eurasia Journal of Mathematics, Science and Technology Education*, *12*(3), 431-455. doi: 10.12973/eurasia.2016.1209a

Gordon, P. R., Rogers, A. M., Comfort, M., Gavula, N., & McGee, B. P. (2001). A taste of problem-based learning increases achievement of urban minority middle-school students. *Educational Horizons, 79*(4), 171-75.

Gresham, F. M., & Gansle, K. A. (1993). Treatment integrity of school-based behavioral intervention studies: 1980-1990. *School Psychology Review*, *22*(2), 254-273.

Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice*, 15, 198-205.

Guilford, J. P. (1950). Creativity. *The American Psychologist, 5*(9), 444-454. doi: http://psycnet.apa.org/

Guilford, J. P. (1959). Three faces of intellect. *American Psychologist, 14*(8), 469-479. doi: 10.1037/h0046827

Guilford, J. P. (1981). Higher-order structure-of-intellect abilities. *Multivariate Behavioral Research, 16*(4), 411-435. doi: 10.1207/s15327906mbr1604_1

Guilford, J. P. (1982). Cognitive psychology's ambiguities: Some suggested remedies. *Psychological Review, 89*(1), 48-59. doi: 10.1037//0033-295X.89.1.48

Guilford, J. P. (1988). Some changes in the structure-of-intellect model. *Educational and Psychological Measurement, 48*(1), doi: 1-4. 10.1177/001316448804800102

Hansen, W. B., Graham, J. W., Wolkenstein, B. H., & Rohrbach, L. A. (1991). Program integrity as a moderator of prevention program effectiveness: Results for fifth-grade students in the adolescent alcohol prevention trial. *Journal of Studies on Alcohol*, 52(6), 568-579. doi: 10.15288/jsa.1991.52.568

Haridza, R., & Irving, K. E. (2017). Developing critical thinking of middle school students using problem based learning 4 core areas (PBL4C) model. *Journal of Physics: Conference Series*, *812*, (1), 1-9. doi: 10.1088/1742-6596/812/1/012081

Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in schools?. *Exceptional Children*, *79*(2), 181-193. doi:10.1177/001440291307900204

Hodges, G. C. (2005). Creativity in education. *English in Education, 39*(3), 47-61. doi:10.1111/j.1754-8845.2005.tb00624.x

Horak, A. K., & Galluzzo, G. R. (2017). Gifted middle school students' achievement and perceptions of science classroom quality during problem-based learning. *Journal of Advanced Academics*, *28*(1), 28-50. doi: 10.1177/1932202X16683424

Hu, W. P., & Adey, P. (2002). A scientific creativity test for secondary school students. *International Journal of Science Education, 24*(4), 389-403. doi:10.1080/09500690110098912

Hyman, R. B. (1978). Creativity in open and traditional classrooms. *The Elementary School Journal, 78*(4), 267-274. doi:10.1086/461111

Inel, D., & Balim, A. G. (2010). The effects of using problem-based learning in science and technology teaching upon students' academic achievement and levels of structuring concepts. In *Asia-Pacific Forum on Science Learning & Teaching*,*11*(2),1-23.

Institute of Education Sciences. (2008). *WWC procedures and standards handbook: Version 2.0*. Washington, DC. Retrieved from http://ies.ed.gov/ncee/ wwc/references/idocviewer/doc.aspx?docid=19&tocid=1

Isaksen, S. G., Dorval, K. B., & Treffinger, D. J. (2011). *Creative approaches to problem solving: A framework for innovation and change*. Thousand Oaks, CA: Sage Publications.

Jo, S., & Ku, J. O. (2011). Problem based learning using real-time data in science education for the gifted. *Gifted Education International*, *27*(3), 263-273. doi: http://gei.sagepub.com

Jo, S., & Maker, C. J. (2011). The effect of the DISCOVER curriculum model on mathematical knowledge and creativity. *Asia-Pacific Journal of Gifted and Talented Education*, *3*(1), 1-17.

Joyce, B., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Kaufman, J. C., Russell, C. M., & Plucker, J. A. (2013). Models and methods of assessing creativity. *Oxford Handbooks Online.* doi:10.1093/oxfordhb/9780199796304.013.0025

Kerr, B. A. (2009). *Encyclopedia of giftedness, creativity, and talent*. Los Angeles, CA; London: SAGE.

Kim, K. (2011). The creativity crisis: The decrease in creative thinking scores on the Torrance tests of creative thinking. *Creativity Research Journal*, 23(4), 285-295. doi: 10.1080/10400419.2011.627805

Kuang, C. C. (2011). Applying the DISCOVER model to nurture creativity in pre-service early childhood teachers. *Gifted Education International*, *27* (3), 300-310. doi:10.1177/026142941102700307

Kuo, C. C., Maker, C. J., Su, F. L., & Hu, C. (2010). Identifying young gifted children and cultivating problem solving abilities and multiple intelligences. *Learning and Individual Differences*, *20* (4), 365-379. doi:10.1016/j.lindif.2010.05.005

Kuo, C., Su, F., & Maker, C. J. (2011). Cultivating problem solving abilities in gifted preschoolers. *Gifted Education International, 27*(3), 311-326. doi: 10.1177/026142941102700308

Kuo, P. H. (2016). Effects of synchronous web-based instruction on students' thinking styles and creativity. *Eurasia Journal of Mathematics, Science & Technology Education*, *12*(3), 609-619. doi: 10.12973/eurasia.2016.1234a

Laerd Statistics (2015). One-way MANOVA using SPSS Statistics. *Statistical tutorials and software guides*. Retrieved from https://statistics.laerd.com/

Little, M. A., Riggs, N. R., Shin, H., Tate, E. B., & Pentz, M. A. (2015). The effects of teacher fidelity of implementation of pathways to health on student outcomes. *Evaluation & the Health Professions, 38*(1), 21-41. doi:10.1177/0163278713489879

Lubart, T. I. (1994). *Product-centered self-evaluation and the creative process*. (Doctoral Dissertation). Available from ProQuest Dissertation and Theses database. (Order No. 9523198). Yale University.

Maker, C. J. (1982). *Curriculum development for the gifted*. Astuin, TX: Pro Ed.

Maker, C. J. (1993). Creativity, intelligence, and problem solving: A definition and design for cross-cultural research and measurement related to giftedness. *Gifted Education International, 9*(2), 68-77. 10.1177/026142949300900202

Maker, C. J. (2001). DISCOVER: Assessing and developing problem solving. *Gifted Education International, 15*(3), 232-251. doi: 10.1177/026142940101500303

Maker, C. J. (2006) Creativity, intelligence, problem-solving and diversity. In: Eriksson, G., & Wallace, B. (Eds.). *Diversity in gifted education: International perspectives on global issues*. New York. Routledge. 28-44.

Maker, C. J. (2017a). *Identifying exceptional talent in STEM*. Manuscript in preparation. Department of Disability and Psychoeducational Studies, University of Arizona, Tucson, Arizona, USA.

Maker, C. J. (2017b). *Culturally responsive assessments of mathematical and spatial analytical skills and abilities: Development, field testing, and implementation*. Manuscript in preparation. Department of Disability and Psychoeducational Studies, University of Arizona, Tucson, Arizona, USA.

Maker, C. J., & Pease, R. (2016). *What do effective teachers do when Implementing the REAPS model to differentiate instruction in a general education classroom?* Manuscript in preparation. Department of Disability and Psychoeducational Studies, University of Arizona, Tucson, Arizona, USA.

Maker, C. J., & Schiever, S. W. (2005). *Teaching models in education of the gifted* (3rd ed.). Austin, TX: Pro-ed.

Maker, C. J., & Schiever, S. W. (2010). *Curriculum development and teaching strategies for gifted learners* (3rd ed.). Austin, TX: Pro-Ed.

Maker, C. J., & Zimmerman, R. (2008). Problem solving in a complex world: Integrating DISCOVER, TASC, and PBL in a teacher education project. *Gifted Education International*, *24*(2-3), 160-178. doi: http://gei.sagepub.com/

Maker, C. J., Jo, S., & Muammar, O. M. (2008). Development of creativity: The influence of varying levels of implementation of the DISCOVER curriculum model, a non-traditional pedagogical approach. *Learning and Individual Differences, 18* (4), 402-417. doi:10.1016/j.lindif.2008.03.003

Maker, C. J., Jo, S., Alfaiz, F. S., & Alhusaini, A. A. (2017). *The test of creative problem solving in science (TCPS-S): Construct and concurrent validity.* Manuscript in preparation. Department of Disability and Psychoeducational Studies, University of Arizona, Tucson, Arizona, USA

Maker, C. J., Muammar, O., Serino, L., Kuang, C. C., Mohamed, A., & Sak, U. (2006). The DISCOVER curriculum model: Nurturing and enhancing creativity in all children. *KEDI Journal of Educational policy*, *3* (2), 99-121.

Maker, C. J., Nielson, A. B., & Rogers, J. A. (1994). Giftedness, diversity, and problem-solving. *Teaching Exceptional Children, 27* (1), 4-19. doi:10.1177/004005999402700101

Maker, C. J., Rogers, J. A., Nielson, A. B., & Bauerle, P. R. (1996). Multiple intelligences, problem solving, and diversity in the general classroom. *Talents and Gifts*, *19*(4), 437-460. doi: 10.1177/016235329601900404

Maker, C. J., Zimmerman, R., Alhusaini, A., & Pease, R. (2015). Real engagement in active problem solving (REAPS): An evidence-based model that meets content, process, product, and learning environment principles recommended for gifted students. *The New Zealand Journal of Gifted Education*, *19*(1). doi: http://www.giftedchildren.org.nz/apex/

Maker, C. J., Zimmerman, R., Pease, R., & Janes, M. (2012). *Real engagement in active problem solving: Differentiation for diverse Learners in regular classrooms.* Unpublished research proposal. University of Arizona.

McMaster, K. L., Jung, P. G., Brandes, D., Pinto, V., Fuchs, D., Kearns, D., ... & Yen, L. (2014). Customizing a research-based reading practice. *The Reading Teacher*, *68*(3), 173-183. doi:10.1002/trtr.1301

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review, 69*(3), 220-232. doi:10.1037/h0048850

Merritt, J., Lee, M. Y., Rillero, P., & Kinach, B. M. (2017). Problem-based learning in K–8 mathematics and science education: A Literature review. *Interdisciplinary Journal of Problem-Based Learning*, *11*(2), 1-14. . doi: http://dx.doi.org/10.7771/1541-5015.1674

Miller, A. L. (2014). A self-report measure of cognitive processes associated with creativity. *Creativity Research Journal, 26*(2), 203-218. doi:10.1080/10400419.2014.901088

Miller, R. B., & Hollist, C. S. (2007). *Encyclopedia of measurement and statistics*. Thousand Oaks, Calif.: SAGE Publications.

Mohamed, A. H. (2006). *Investigating the scientific creativity of fifth-grade students* (Doctoral Dissertation). Available from ProQuest Dissertation and Theses database. University of Arizona.

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*(3), 247-266. 10.1016/0272-7358(91)90103-2

Mowbray, C, Holter, M. C, Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, *24*, 315-340. doi:10.1016/S1098-2140(03)00057-2

Mumford, M. D., & Gustafson, S. B. (1988). Creativity syndrome: Integration, application, and innovation. *Psychological Bulletin, 103*(1), 27-43. doi:10.1037/0033-2909.103.1.27

Mundilarto, H. I. (2017). Effect of problem-based learning on improvement physics achievement and critical thinking of senior high school student. *Journal of Baltic Science Education*, *16*(5). 761-779.

National Advisory Committee on Creative and Cultural Education (1999). Retrieved from http://sirkenrobinson.com/pdf/allourfutures.pdf

National Research Council. (2004). *On evaluating curricular effectiveness: Judging the quality of K-12 mathematics evaluations*. National Academies Press.

Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services & Research*, *39*(4), 374-396. doi:10.1007/s11414-012-9295-x

No Child Left Behind Act, Title IX, General Provisions, Part A, Section 9101(2001). Retrieved December 20, 2006, from http://www.ed .gov/policy/elsec/leg/esea02/pg 1 07.html

Nurdin, S., & Setiawan, W. (2015). Improving students' cognitive abilities and creative Thinking skills on temperature and heat concepts through an exelearning-assisted problem based learning. *International Journal of Scientific & Technology Research*, *4*(8), 59-63.

Nurdyani, F., Slamet, I., & Sujadi, I. (2018). Creative thinking level of students with high capability in relations and functions by problem-based learning. *Journal of Physics: Conference Series,* 983, (1), 1-5. doi: 10.1088/1742-6596/983/1/012102

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*(1), 33-84. doi:10.3102/0034654307313793

Olson, C., & Hernstein, R .J. (1976). On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, 83(4), 579-586.

Patterson, G., & Graham, W. (2018). *Clinical interventions in criminal justice settings*. New York: Academic Press.

Problem. (n.d.), In Merriam-Webster's online dictionary (11th ed.). Retrieved from http://www.m-w.com/dictionary/problem

Przybylski, R., & Orchowsky, S. (2015). *Implementing evidence-based practices*. Washington, DC: Justice Research and Statistics Association, Bureau of Justice Statistics, U.S. Department of Justice, National Criminal Justice Association.

Ratnasari, D., Supriyanti, T., & Rosbiono, M. (2017). Vocational high school students' creativity in food additives with problem based learning approach. *Journal of Physics: Conference Series*, *895*(1), 1-6. doi: 10.1088/1742-6596/895/1/012001

Reinoso, J. (2011). Real-life problem solving: Examining the effects of alcohol within a community on the Navajo nation. *Gifted Education International, 27*(3), 288-300. doi: 10.1177/026142941102700306

Rhodes, M. (1961). An analysis of creativity. *Phi Delta Kappan*, 42, 305–310. doi: http://www.jstor.org/

Riley, T., Webber, M., & Sylva, K. (2017). Real engagement in active problem solving for Māori boys: A case study in a New Zealand secondary school. *Gifted and Talented International*, 32(2), 75-86. doi: https://doi.org/10.1080/15332276.2018.1522240

Ringwalt, C. L., Ennett, S., Johnson, R., Rohrbach, L. A., Simons-Rudolph, A., Vincus, A., & Thorne, J. (2003). Factors associated with fidelity to substance use prevention curriculum guides in the nation's middle schools. *Health Education & Behavior*, *30*(3), 375-391. doi:10.1177/1090198103253627

Rogers, E. M. (1983). *Diffusion of innovations* (3rd ed.). New York: Free Press.

Saccardi, M. (2014). *Creativity and children's literature: New ways to encourage divergent thinking*. Denver, CO: Libraries Unlimited.

Sak, U., & Maker, C. (2006). Developmental variation in children's creative mathematical thinking as a function of schooling, age, and knowledge. *Creativity Research Journal,18*(3), 279-291.

Salemi, M. L. (2010). *Utilizing fluency, flexibility, originality, and elaboration to enhance creativity and vocabulary use for improving reading comprehension in third through sixth grade students*. (Doctoral Dissertation). Available from ProQuest Dissertation and Theses database. (Order No. 3443006). Union University.

Schoenwald, S. K., Sheidow, A. J., & Letourneau, E. J. (2004). Toward effective quality assurance in evidence-based practice: Links between expert consultation, therapist fidelity, and child outcomes. *Journal of Clinical Child and Adolescent Psychology*, *33*(1), 94-104. doi:10.1207/S15374424JCCP3301_10

Scratchley, L. S., & Hakstian, A. R. (2000). The measurement and prediction of managerial creativity. *Creativity Research Journal, 13*(3-4), 367-384. doi:10.1207/S15326934CRJ1334_14

Seidman, I. (2013). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. (4th Ed). New York: Teachers college press.

Shively, C. H. (2011). Grow creativity! focusing on fluency, flexibility, originality, and elaboration skills gives teachers and students an effective shortcut to developing creativity together. *Learning & Leading with Technology, 38*(7), 10.

Sholomskas, D. E., Syracuse-Siewert, G., Rounsaville, B. J., Ball, S. A., Nuro, K. F., & Carroll, K. M. (2005). We don't train in vain: a dissemination trial of three strategies of training clinicians in cognitive-behavioral therapy. *Journal of consulting and clinical psychology*, *73*(1), 106-115. doi:10.1037/0022-006X.73.1.106

Siew, N. M., & Chin, M. K. (2018). Design, development and evaluation of a problem-based with cooperative module on scientific creativity of preschoolers. *Journal of Baltic Science Education*, *17*(2). 215-228.

Siew, N. M., & Mapeala, R. (2016). The effects of problem-based learning with thinking maps on fifth graders' science critical thinking. *Journal of Baltic Science Education*, *15*(5). 602-616.

Siew, N. M., Chin, M. K., & Sombuling, A. (2017). The effects of problem based learning with cooperative learning on preschoolers' scientific creativity. *Journal of Baltic Science Education*, *16*(1).100-112.

Siew, N. M., Chong, C. L., & Lee, B. N. (2015). Fostering fifth graders' scientific creativity through problem-based learning. *Journal of Baltic Science Education*, *14*(5), 655-669.

Silvia, P. J., Wigert, B., Reiter-Palmon, R., & Kaufman, J. C. (2012). Assessing creativity with self-report scales: A review and empirical evaluation. *Psychology of Aesthetics, Creativity, and the Arts, 6*(1)*,* 19-34. doi:10.1037/a0024071

Smith, J. D., Schneider, B. H., Smith, P. K., & Ananiadou, K. (2004). The effectiveness of whole-school antibullying programs: A synthesis of evaluation research. School Psychology Review, *33*, 547–560.

Steinbright Career Development Center (2018). *Types of interviews*. Retrieved from http://drexel.edu/scdc/professional-pointers/interviewing/interview-types/

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge: Cambridge University Press.

Sternberg, R. J. (1999). Intelligence as developing expertise. *Contemporary Educational Psychology, 24*(4), 359-375. doi:10.1006/ceps.1998.0998

Sternberg, R. J. (2005). The theory of successful intelligence. *Interamerican Journal of Psychology, 39*(2), 189-202.

Sungur, S., Tekkaya, C., & Geban, Ö. (2006). Improving achievement through problem-based learning. *Journal of Biological Education*, *40*(4), 155-160. doi: 10.1080/00219266.2006.9656037

Swanson, E., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L. (2013). Intervention fidelity in special and general education research journals. *The Journal of Special Education*, *47*(1), 3-13. doi:10.1177/0022466911419516

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th Ed). Upper Saddle River, New Jersey: Pearson Education, Inc.

Tan, O. S. (2003). Problem-based learning innovation: Using problems to power learning in the 21st century. Thomson Learning Asia.

Thomas, N. G., & Berk, L. E. (1981). Effects of school environments on the development of young children's creativity. *Child Development, 52*(4), 1153-1162. doi:10.1111/j.1467-8624.1981.tb03161.x

Tobler, N. S. (1986). Meta-analysis of 143 adolescent drug prevention programs: Quantitative outcome results of program participants compared to a control or comparison group. *Journal of Drug Issues*, 16, 537–567. doi:10.1177/002204268601600405

Tomlinson, C. A. (2001). *How to differentiate instruction in mixed-ability classrooms*. (2nd ED). Alexandria, VA: Association for Supervision & Curriculum Development.

Torrance, E. P. (1963). *Education and the creative potential*. Minneapolis: University of Minnesota Press.

Torrance, E. P. (1968). A longitudinal examination of the fourth grade slump in creativity. *Gifted Child Quarterly, 12*(4), 195-199. doi:10.1177/001698626801200401

Torrance, E. P. (1977). Creativity in the classroom. Washington, DC: National Education Association.

Torrance, E. P. (1988). The nature of creativity as manifest in its testing. In R. J. Sternberg (Ed.), *he Nature of creativity : Contemporary psychological perspectives*. (pp.43-75). New York: Cambridge University Press.

Torrance, E. P. (1993). Understanding creativity: Where to start? *Psychological Inquiry, 4*(3), 232-234. doi:10.1207/s15327965pli0403_17

Torrance, E. P., & Shaughnessy, M. F. (1998). An interview with E. Paul Torrance: About creativity. *Educational Psychology Review, 10*(4), 441-452. doi:10.1023/A:1022849603713

Tracy, B. (2015). *Creativity and problem solving*. New York, N.Y: AMACOM.

Urban, K. K. (1991). On the development of creativity in children. *Creativity Research Journal*, *4*(2), 177-191. doi: 10.1080/10400419109534384

Urdang, L. (1993). *The Oxford thesaurus: an AZ dictionary of synonyms*. New York: Oxford University Press.

Vincent, A. S., Decker, B. P., & Mumford, M. D. (2002). Divergent thinking, intelligence, and expertise: A test of alternative models. *Creativity Research Journal, 14*(2), 163-178. doi:10.1207/S15326934CRJ1402_4

Wallace, B., Bernardelli, A., Molyneux, C., & Farrell, C. (2012). TASC: Thinking actively in a social context. A universal problem-solving process: A powerful tool to promote differentiated learning experiences. *Gifted Education International, 28*(1), 58-83. doi: 10.1177/0261429411427645

Wallace, Belle. (2008). The early seedbed of the growth of TASC: Thinking actively in a social context. *Gifted Education International*, 24(2-3), 139-156. doi: 10.1177/026142940802400303

Wartono, W., Diantoro, M., & Bartlolona, J. R. (2018). Influence of problem based learning model on student creative thinking on elasticity topics a material. *Jurnal Pendidikan Fisika Indonesia*, *14*(1), 32-39. doi: 10.15294/jpfi.v14i1.10654

Webber, M., Riley, T., Sylva, K., & Scobie-Jennings, E. (2018). The Ruamano project: Raising expectations, realising community aspirations and recognising gifted potential in Mori boys. *Australian Journal of Indigenous Education,* 1-12. doi.org/10.1017/jie.2018.16

Welsch, P. K. (1980). *The nurturance of creative behavior in educational environments: A comprehensive curriculum approach (*Doctoral Dissertation). Retrieved from University of Michigan.

Welsh, B. C., Sullivan, C. J., & Olds, D. L. (2010). When early crime prevention goes to scale: A new look at the evidence. *Prevention Science*, *11*(2), 115–125. doi:10.1007/s11121-009-0159-4

West, H. (2008). A study of the impact of using the TASC wheel on children's involvement in their own learning. *Gifted Education International*, *24*(2-3), 297-304. doi: http://gei.sagepub.com/

Willmoth, A. (2008). Using TASC to structure a maths investigation. *Gifted Education International*, *24*(2-3), 246-248. doi: http://gei.sagepub.com/

Wilson, F. S., Stuckey, T., & Langevin, R. (1972). Are pupils in the open plan school different? *The Journal of Educational Research, 66*(3), 115-118. doi:10.1080/00220671.1972.10884422

Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta analysis. *Journal of Consulting and Clinical Psychology*, 71, 136–149. doi:10.1037/0022-006X.71.1.136

Wong, K. K. H., & Day, J. R. (2009). A comparative study of problem-based and lecture-based learning in junior secondary school science. *Research in Science Education*, *39*(5), 625-642. doi: 10.1007/s11165-008-9096-7

Wu, I., Pease, R., & Maker, C. J. (2015). Students' perceptions of real engagement in active problem solving. *Gifted and Talented International, 30*(1-2), 106-121. doi:10.1080/15332276.2015.1137462

Yew, E. H., & Goh, K. (2016). Problem-based learning: an overview of its process and impact on learning. *Health Professions Education*, *2*(2), 75-79. doi: 10.1016/j.hpe.2016.01.004

Yulindar, A., Setiawan, A., & Liliawati, W. (2018). Enhancement of problem solving ability of high school students through learning with real engagement in active problem solving (REAPS) model on the concept of heat transfer. *Journal of Physics*, *1013*, (1), 1-7.

Zeng, L., Proctor, R. W., & Salvendy, G. (2009). Fostering creativity in service development: Facilitating service innovation by the creative cognition approach. *Service Science*, *1*(3), 142-153. doi: org/doi/abs/10.1287/serv.1.3.142

Zhang, M., Parker, J., Eberhardt, J., & Passalacqua, S. (2011). "What's so terrible about swallowing an apple seed?" Problem-based learning in kindergarten. *Journal of Science Education and Technology*, *20*(5), 468-481. doi: 468. 10.1007/s10956-011-9309-0

Zimmerman, R., Maker, C. J., Gomez-Arizaga, M. P., & Pease, R. (2011). The use of concept maps in facilitating problem solving in earth science. *Gifted Education International, 27*(3), 274-287. doi:10.1177/026142941102700305

Zvoch, K. (2009). Treatment fidelity in multisite evaluation: A multilevel longitudinal examination of provider adherence status and change. *American Journal of Evaluation,* 30(1), 44-61. doi:10.1177/1098214008329523